ALPHA FOUNDATION FOR THE IMPROVEMENT OF MINE SAFETY AND HEALTH

Revised Final Technical Report

Title: Improving Communication in Noise for Miners Wearing Hearing Protection: Algorithms for Mine Machinery Noise

Grant Number: AFC518SP-92

Organization Name: University of Connecticut Health

Principal Investigator: Anthony J. Brammer

Contact Information: brammer@uchc.edu; Tel. (860) 679-2760; FAX. (860) 679-1989

Period of Performance: January 01, 2020 - November 30, 2021

This study was sponsored by the Alpha Foundation for the Improvement of Mine Safety and Health, Inc. (ALPHA FOUNDATION). The views, opinions, and recommendations expressed herein are solely those of the authors and do not imply any endorsement by the ALPHA FOUNDATION, its Directors and staff.

1.0 Executive Summary

In 2019 we completed an Alpha Foundation Proof-of-Concept Technology Development project entitled "Improving Communication in Noise for Miners Wearing Hearing Protection" (AFC518-10). As a result of this study, the Foundation contracted that we undertake *"further proof-of-concept development such that it can reach a level whereby sufficient confidence can be ascertained to justify continued advancement to a functional prototype that could be demonstrated in an actual mine or high-fidelity simulated mine environment"*. In particular, refinement of the algorithms was required for "*different environmental noises, different sound levels and different signal to noise ratios to determine the capacity to preserve speech intelligibility in a broader range of conditions particularly those that are emblematic of the frequency spectrum of typical mine noises (especially from mine machinery that may be intermittent and variable in intensity)".* These statements define our mission for the present work.

In our previous study for the Foundation, we developed a family of algorithms designed to improve the intelligibility of speech. All commenced by dividing the input sounds into separate, contiguous frequency "subbands" using a set of band-pass filters arranged in parallel. Within each subband there are separate signal paths and control paths that operate in parallel. The former contains the components of speech and environmental noise within the bandwidth of the subband, while the latter contains the signal processing we design to improve speech understanding by modifying the former.

In our initial study the results obtained by directly changing, or modulating, the amplitude of the sounds in the signal path were encouraging. An alternative to the *linear* modulation for controlling the amplitude of signals in individual subbands used in direct modulation (DM) is to employ *binary* modulation in which the gain is either switched "on" or "off" (commonly described as applying a binary mask (BM)). The method has been reported extensively in the scientific literature for circumstances in which speech and environmental noise can be obtained separately, in which case the processing is described as applying an ideal binary mask (IBM). This situation does not occur during face-to-face communication in mining (speech and noise are always intermixed and never available separately). However, it could occur when listening in a mining environment to a remote talker over a wireless or wired link using a communication headset or electronic hearing protection device (eHPD). Thus a study of both an IBM and a BM provides valuable insights into the potential benefits of applying binary masking to improve communication in mines.

The assessment of the performance of algorithms developed in this study was by listening tests. Now, conventional listening tests require subjects to come to our facilities on campus. However, many volunteers were unwilling to come to the university during the COVID-19 pandemic. For this reason we have developed and validated a web-based listening test for subjects to undergo at home, or elsewhere, and demonstrated that it is an adequate alternative to performing tests on campus for persons with normal hearing.

The performance of algorithms involving direct linear modulation (DM) and binary modulation both with and without access to speech and noise separately (i.e., IBM and BM algorithms, respectively), has been established in different noises and at different speech signal-to-noise ratios (SNRs). The noise of a continuous miner and roof bolter were selected to represent mining noises, and a generalized industrial-like noise was also used in the development of algorithms. The noises possessed different frequency spectra that enabled the performance of the algorithms to be determined in noise environments believed to be typical of those found in mines.

The use of different SNRs in the listening tests implies that they were performed at different sound levels. Moreover, the conduct of listening tests under three different experimental configurations - one conducted under controlled conditions within a controlled environment supervised by trained audiologists, a second conducted in an audiometric room in our laboratory in which subjects themselves fitted their earphones with assistance if necessary, and a third conducted commonly at home where subjects chose the headphones or earphones to wear and the sound level for the tests - also introduced an uncontrolled range of sound levels into each test.

Under all the operating conditions imposed by the listening tests, including continuous and intermittent noises, our IBM algorithm improved speech intelligibility over that without signal processing. The results included large increases in word scores under the most difficult listening conditions (up to 37%), when the noise was loudest, and no reduction in intelligibility when there was little or no noise to compromise intelligibility. The benefit to a user of this technology occurred for all conditions we evaluated listening to speech in noise. When the unprocessed word score is high, say 80 - 90% words correct, little improvement in intelligibility is required to aid conversation, and so the processed word score is only slightly greater than the unprocessed score. Conversely, when the unprocessed word score is low, say 30 - 40% words correct, substantial improvement in intelligibility is required to aid the listener understand the speech, and the IBM algorithm delivers word scores in the range of 70% words correct. This is a marked improvement that can be expected to greatly influence communication and workers' safety. Consequently we believe this algorithm can be used effectively under all listening conditions in noise to improve speech intelligibility, and has demonstrated the functional capability for in-service operational application to situations in which speech and noise are available separately.

The DM and 24-subband BM algorithms produced smaller improvements in speech intelligibility compared to that without signal processing than the IBM. Both algorithms improved intelligibility under almost all listening conditions and all noises used, with word scores increasing by up to 10%. While smaller improvements were expected for both algorithms, they nevertheless confirm the potential for improving speech intelligibility by our signal processing during face-to-face communication, when speech and noise are intermixed and never available separately. Implementation of either of these algorithms in a communication headset or an eHPD would immediately provide modest improvements in intelligibility, but both DM and BM would benefit from further refinement. Based on the work reported here and the algorithms evaluated in our previous study for the Alpha Foundation, it is not clear how the DM algorithm could be modified to improve its performance. However, it should be possible to improve the performance of the BM algorithm to approach that of our IBM algorithm described above. Accordingly, the limitations of the BM algorithm are being addressed after the completion of this study and progress is described in an Appendix.

Even after an algorithm is developed that can increase the intelligibility of speech "buried" in noise, it must be transferred to electronics capable of microminiaturization. The computational complexity of 24-subband algorithms will require careful implementation to function throughout a work shift within a small, lightweight package suitable to be worn as part of a miner's equipment or attached to, or integrated into, a miner's helmet. We judge the Technology Readiness Level of the proof-of-concept to be level TRL 3 ("Analytical and experimental critical function and/or characteristic proof of concept" - NASA usage, and "Experimental proof of concept" - European Union usage of the current nine-unit scale).

2.0 Technology Description and Mission Statement

2.1 Mission Statement

In January 2019, we completed an Alpha Foundation Proof-of-Concept Technology Development project in the focus area of Advanced Personal Protective Equipment entitled "Improving Communication in Noise for Miners Wearing Hearing Protection" (AFC518-10). Our mission was to develop a method for improving communication in noise suitable for miners wearing a hearing protection device (HPD), in order to reduce confusion identifying spoken words and increase the audibility of warning sounds. These improvements should reduce the risk of miners being struck by moving equipment and errors in speech communication between co-workers.

We noted in our *Final Technical Report* that the performance of our algorithms remains to be established for different talkers, different environmental noises, different sound levels, and different speech signal to noise ratios (SNRs). These limitations of the original work were recognized by the Alpha Foundation project review. In consequence, the Foundation contracted that we undertake *"further proof-of-concept development such that it can reach a level whereby sufficient confidence can be ascertained to justify continued advancement to a functional prototype that could be demonstrated in an actual mine or high-fidelity simulated mine environment"*. In particular, refinement of the algorithms is required for "*different environmental noises, different sound levels and different SNRs to determine if the capacity to preserve speech intelligibility in a broader range of conditions particularly those that are emblematic of the frequency spectrum of typical mine noises (especially from mine machinery that may be intermittent and variable in intensity)".* The additional proof-of-concept required by the Alpha Foundation's review of our original study defines our mission statement for the present study.

The end goal of the technology being developed is to implement the algorithms within a small, lightweight electronics package that could function effectively throughout a work shift and be worn as part of a miner's equipment or attached to, or integrated into, a miner's helmet. It is envisaged that the device would otherwise function much like current-day sound level dependent electronic hearing protectors, which have been gaining popularity in recent years.

2.2 Health and Safety Mining Need

The US mining industry has the highest prevalence of hazardous workplace noise exposures of all industrial sectors (Tak et al., 2009). According to the National Institute of Occupational Safety and Health (NIOSH), one in four miners have a hearing problem and, by retirement age, four out of five mine workers have impaired hearing. The unwillingness of workers to wear commercially available HPDs because of fear they will not be able to understand co-workers speech or hear warning sounds has been repeatedly documented in the literature (for reviews, see Suter, 1992; Suter 2001). This contributes to the avoidance of hearing protector use by up to 50 % of some noise-exposed worker groups (McKinley et al., 2005; Morata et al., 2001). In miners' focus group sessions, the priority of underground survival was ranked well above the "nuisance" of hearing loss (Murray-Johnson et al., 2004; Patel et al., 2001). As succinctly stated by Azman and Hudak (2011), "*miners often complain of reduced audibility or confusion identifying spoken words when wearing conventional hearing protectors. This leads to an increased risk of miners being struck by moving equipment or errors in communication with co-workers*".

Failure to hear environmental and warning sounds is an additional concern for job safety for miners with subclinical hearing loss (Morata et al., 2005), which compromises audibility and has long been associated with increased risk of injury in a noisy workplace (for review, see Wilkins et al., 1987). In a study focusing on hearing acuity, noise and hearing loss accounted for more than 40% of the injuries occurring in a shipyard (van Charante et al., 1990). The elevated risk of injury when wearing existing commercial HPDs even for persons with normal hearing has also been documented (Choi et al., 2005). In their study of agricultural injuries, the relative risk of injuries to workers wearing HPDs was 2.2, and was independent of their hearing acuity. NIOSH's National Traumatic Occupational Facility Surveillance System records 204 accidental deaths of pedestrians in industry struck by forklifts from 1980 to 1994 (Collins et al., 1999). While the causes of these accidents cannot be deduced, a Fatality Assessment and Control Evaluation (FACE) report of a worker wearing an HPD, who died after being run over by a log loader reversing with its back-up alarm sounding, would appear to be an example of the failure to identify the warning sound (Anon, 1995).

NIOSH in its *Criteria for a Recommended Standard: Occupational Exposure to Noise* identified the "persistent problems" of HPDs, concluding that "*Research should also lead to the development of hearing protectors that eliminate troublesome barriers by . . . improved speech intelligibility and audibility of warning signals"* (Anon, 1998, p. 71). In this study, we have focused on reducing communication problems when the talker is in the same environment as a listener, and consequently on methods for improving the intelligibility of face-to-face speech communication suitable for users of hearing protection. A secondary consideration has been situations in which the talker communicates with a listener over a wireless or wired link, and is not subjected to the intensity of noise experienced by the listener. A successful method, or methods, would enable the development of improved HPDs that incorporate the appropriate electronic processing of sounds (eHPDs). Such devices could ultimately lead to greater acceptability, and consequently wider use, of eHPDs in mines, hence reducing the risk of noise-induced hearing loss and accidents.

3.0 Technology Description and Design Strategy

3.1 Shortcomings of Previous Technology Approaches

While there have been numerous attempts to reduce the noise of machinery used in mechanized mining, it is generally recognized that many miners remain potentially overexposed to noise (Babich and Bauer, 2006; Joy and Middendorf, 2007). During the last twenty years, specialized hearing protectors have been developed for situations in which noise levels change in space or with time, such as when walking towards or around a machine or when a nearby vehicle moves or stops operating. In these circumstances, the device automatically adjusts the amount of hearing protection to enable the user to hear more sounds in the environment when there is less environmental noise. These so-called sound level dependent HPDs (sometimes called level dependent HPDs, or sound restoration HPDs) are gaining popularity, and their applicability to mining environments has been studied (Azman & Hudak, 2011). Several have been approved by the Mines Safety and Health Administration for use underground (see reference to website). They employ electronically-controlled sound transmission from the environment surrounding the user to the ear, and for this purpose include a microphone outside the eHPD, processing electronics, and a miniature earphone or loudspeaker located in the ear cup or earplug. For listening to remote talkers, there is also a wireless or wired link in some eHPDs. While the technology can improve face-to-face communication when the noise levels are low, current devices fail to improve the intelligibility of speech in the presence of loud noises, such as those occurring in underground mines (e.g., from continuous mining machines, or roof bolting machines), compared to when conventional hearing protectors are worn (Azman & Hudak, 2011). The same conclusion has been drawn in other studies using different environmental noises (Dolan & O'Loughlin, 2005; Plyler & Klumpp, 2003), though not in a survey of HPD preferences in an industrial setting (Tufts et al., 2011). When the environmental noise is sufficiently loud, a level dependent HPD is designed to cut-off electronically all sounds from outside the eHPD, so speech from a nearby talker as well as the noise will not be heard. The methods developed here are intended to improve communication in all situations.

There have been several attempts reported in the literature to improve face-to-face communication in a noisy environment, mostly intended for application to hearing aids. A recent study has described a method for reducing noise (as opposed to improving intelligibility) when wearing eHPDs and attempting to communicate face-to-face in a noisy environment (Lezzoum et al., 2016). The method involved first dividing the frequency spectrum of the sounds (i.e., the combined speech plus environmental noise) into narrow bands of frequencies, commonly termed "subbands". The instantaneous magnitude of the envelope of the signal in each subband was then used to control the instantaneous gain applied to the signal in that subband. The subbands signals were then recombined and the processed sounds formed an audio signal whose properties could be evaluated or presented to a listener. The complete process was ongoing, leading to a time-varying gain in each subband based on the envelope of the sound pressure in that subband. Lessoum et al. (2016) observed that listeners reported hearing reduced noise and improved sound *quality* when the speech was initially mostly intelligible.

A method for simultaneously reducing the environmental noise at the ear and improving the intelligibility of speech from a remote talker when the listener is wearing an eHPD has been described by Brammer et al. (2014). Substantial improvements in speech intelligibility were obtained using subband active noise control (SANC) to increase the speech SNR at the ear. Thus, SANC provides an alternate approach for producing time-varying subband gains, but has received little attention in the literature for improving communication.

For stand-alone devices (i.e., devices without access to remote computational resources), two methods for changing the gain have been commonly reported in the literature. The first, as already described, involves *direct modulation* (DM) (Apoux et al., 2004; Chung et al., 2009; Clarkson and Bahgat, 1991; Langhans and Strube, 1982; Lezzoum et al., 2016; Lorenzi et al., 1999; van Buuren et al., 1999; Wiinberg et al., 2018), and the second involves *binary modulation* or, as it is frequently called, *binary masking* (BM). In the latter method, the control signal switches the gain applied to a given subband on or off, depending on whether the sounds in the subband contain mostly those desired to be heard (switch "on") or undesired noise (switch "off").

Overall, the results obtained in studies approximating "real-world" conditions have been inconsistent, with some finding a small improvement in intelligibility under some conditions of speech SNR (Clarkson and Bahgat, 1991; Lorenzi et al., 1999; Wiinberg et al., 2018), up to ~10% improvement for some SNRs and noises (Apoux et al., 2004; Chung et al., 2009), and other studies finding no improvement (Clarkson and Bahgat, 1991; Langhans and Strube, 1982, van Buuren et al., 1999). When speech and noise are in separate environments, such as when listening to a remote talker over a radio link, substantial improvements in intelligibility can be obtained, reaching as much as ~50% with undisclosed computational resources (Anzalone et al., 2006; Arehart et al., 2015; Brungart et al., 2006; Kim et al., 2009; Kjems et al., 2009; LI and Loizou, 2008; Wójcicki and Loizou, 2012).

Our challenge is to approach the best performance recorded elsewhere with algorithms that can be implemented in a small, stand-alone electronic package that could be worn by a miner throughout a work shift.

3.2 Rationale and Design Strategy for the Proposed Approach

In past work for the Alpha Foundation, we have developed a family of algorithms designed to improve the intelligibility of speech in noise (see *Algorithm Development* in the Revised Final Technical Report to Grant AFC518-10, 2019). All our algorithms commence by dividing the input sounds into separate, contiguous frequency subbands using a set of band-pass filters arranged in parallel. The frequency range chosen is from 200 Hz to 6 kHz. As previously described, within each subband there is a signal path and a control path that operate in parallel. The signal path contains the components of speech (or warning sounds) and environmental noise with frequencies that fall within the range of the corresponding band-pass filter. The control path contains signal processing designed to improve speech understanding, and first forms the envelope of the sounds in each subband. It is the envelope waveform that we use to linearly amplitude modulate the speech in noise contained in the signal path.

Our initial results using direct modulation were encouraging, and for this reason we have continued to refine the method. We have also considered an alternative to linear modulation for controlling the gain of signals in individual subbands and for this purpose have explored binary modulation. The method has been reported extensively in the literature for circumstances in which speech and environmental noise can be obtained separately, in which case it is termed an *ideal binary mask* (IBM). This situation does not occur during face-to-face communication in mining (speech and noise are always intermixed and never available separately). However, it could occur when listening in a mining environment over a wireless or wired link to a remote speaker using a communication headset or eHPD. Thus a study of both an IBM and a BM provides valuable insights into the potential benefits of applying binary modulation to improve communication in mining environments.

A concept block diagram for improving communication using an IBM is shown in Figure 1. As in previous algorithms, the signals are processed in subbands, each of which contains a band-pass filter that collectively span the frequency range in which speech (and warning sounds) occur. The frequency range chosen is from 200 Hz to 6 kHz, as before.

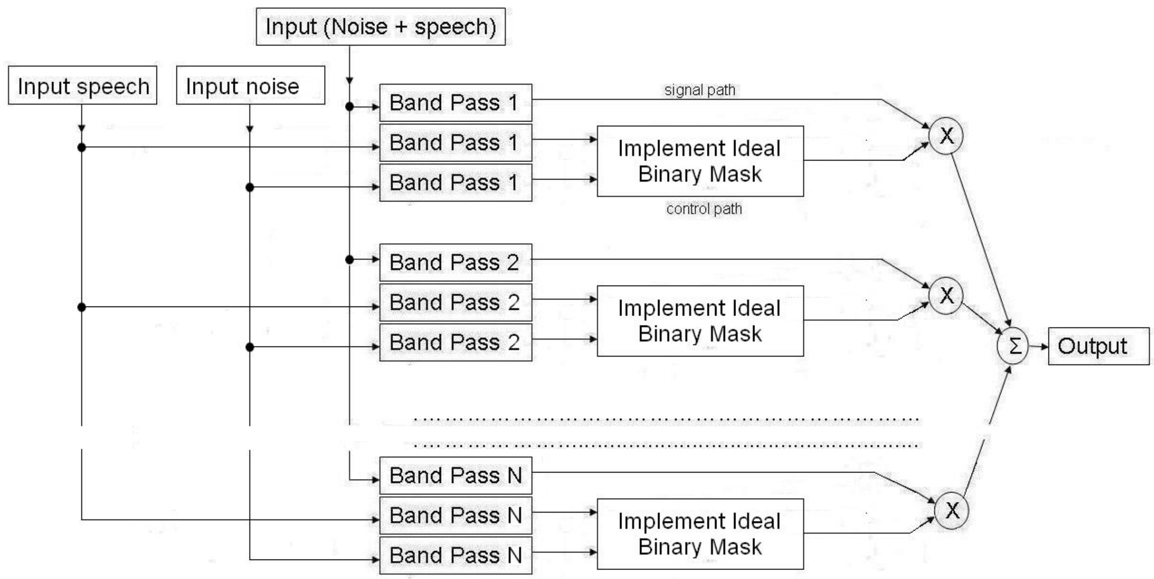


Figure 1: Concept block diagram for computing an Ideal Binary Mask (IBM) for speech in noise. The speech in noise, which is inputted to the signal path, is also available at separate inputs as speech, and noise. The speech and noise are inputted separately to the control path for computation of the IBM in each subband.

There are now, however, three inputs to the algorithm: speech alone, noise alone, and the same speech mixed with the same noise. As in our other algorithms for improving speech intelligibility, the speech in environmental noise forms the input to the signal path, while the speech and environmental noise now form separate inputs to the control path. The aim of any binary mask is to determine whether the signal path at a particular time contains mostly speech or mostly noise. This is done in an IBM by forming the SNR, which is possible in this case as both the speech and noise signals are known separately.

By introducing a threshold value for the SNR, it is possible to segregate the speech in noise in the signal path into times when speech (or warning) sounds dominate (i.e., the SNR is greater than the threshold value), and times when noise dominates (i.e., the SNR is less than the threshold value). The mask is then implemented by multiplying the signal path by unity when the SNR is greater than the threshold and zero when the SNR is less than the threshold. In this way, time segments of speech in noise are either passed unchanged through a subband to be summed with the outputs of other subbands (shown by the "Σ" in Figure 1), or eliminated. Hence a processed signal is produced forming the output of the algorithm that may be recorded and evaluated by listeners. Clearly, the binary nature of the process introduces frequency gaps in the sounds, so that in situations in which noise dominates in most subbands and most SNRs are less than the threshold, little sound would reach the output of an IBM.

An immediate question surrounds the consequences of introducing "holes" in the frequency spectrum of speech on its intelligibility. In unrelated studies it has been demonstrated that considerable loss of frequency content in separate frequency bands produces little reduction in speech understanding (Warren et al., 1995), the brain apparently being capable of "filling in" the spectral gaps. In other words, there is considerable redundancy in speech sounds, a redundancy that we are attempting to exploit by using an IBM or BM. This observation, however, also implies that not all changes in the physical characteristics of speech in noise will necessarily result in changes in speech intelligibility.

With this concern resolved, it is next necessary to select the number of subbands. Previous published work has commonly used a large number of narrow subbands (e.g., 64) (Arehart et al., 2015), which has the consequence of potentially introducing only small "holes" in the frequency spectrum if isolated subbands are masked (i.e., gain is zero). While this approach

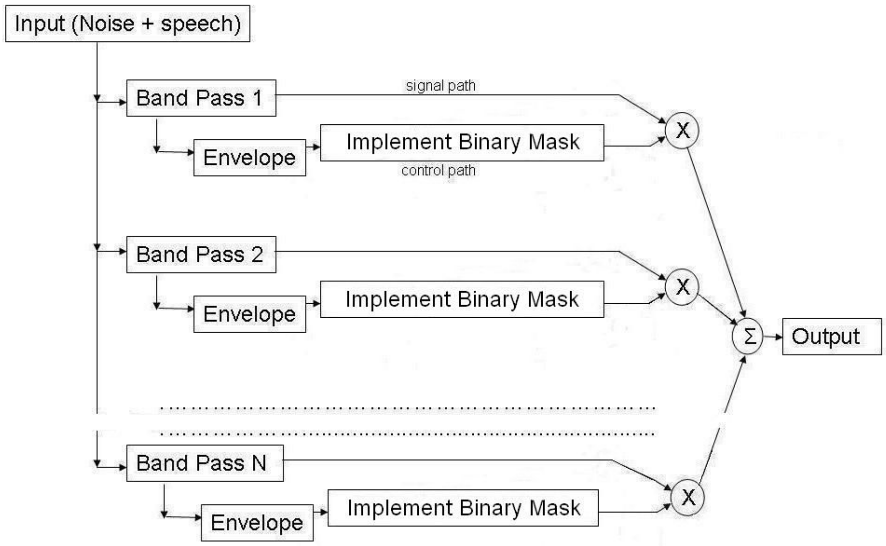


Figure 2: Concept block diagram for computing a Binary Mask (BM) for speech in noise. The speech in noise is inputted to the signal path. The envelope of this signal is used in the control path to compute the Binary Mask in each of N subbands (N = 24).

has logical appeal, it would be impractical for our study, which is focused on developing a method that could be applied today to an eHPD. Moreover, as already mentioned, there is not a one-to-one correlation between physical changes to speech signals and changes to intelligibility.

A basis for the number of subbands to employ may be better informed by the psychology of hearing. It is well known that the ear responds differently to combinations of sounds and noise depending on their frequencies and bandwidths, and the process of masking sounds by noise (in which the noise prevents the sound from being heard) depends on the so-called critical bandwidth (Moore, 2013). It is found generally that noise with bandwidth equal to, or less than, a critical band can only mask sounds at frequencies within this frequency band. Hence, choosing subbands with bandwidth in excess of the critical bandwidth will result in the noise in a subband being able to mask sounds in more than one critical band, thereby rendering an IBM inefficient (viz.: if the IBM sets a subband gain of zero, will there be unnecessary excess loss of speech information?). A solution to the problem of selecting the minimum number of subbands consistent with an efficient IBM is therefore to set the subband bandwidth to the critical bandwidth. Published values for critical bandwidths at different frequencies lead us to employ 24 subbands for sounds in the frequency range from 200 Hz to 6 kHz (Moore, 2013). Accordingly, a 24-subband algorithm has been created to implement an IBM. Note that previous algorithms contained 16 subbands.

Now it is evident that a BM, rather than an IBM, will have to be developed for situations in which speech and noise are premixed and not available separately, such as for our application to face-to-face communication between workers in mines. This is not without challenge as to the best of our knowledge no successful method for constructing a BM has been reported in the literature. The approach we propose to develop is based on our observation, previously reported, that the frequency spectra of the envelopes of waveforms for speech, tonal warning sounds and environmental noises are distinctly different, and hence provides a means to identify the presence of each of these sounds. The algorithm is shown in concept in Figure 2. It contains almost all the elements of a DM algorithm except in the control path, where the "implement modulation" operation is replaced by "implement binary mask".

The process of modifying the speech (or warning sound) plus environmental noise in a subband is shown by the "X" in Figure 2. It involves multiplying the contents of the signal path by the output of the binary mask in the control path, which, as for the IBM, is either, zero or unity. Finally, the modified signals from each subband are combined (shown by the Σ in Figure 2) and the process is repeated. In this way, as in our other algorithms, a processed output signal is produced that may be recorded and subsequently replayed for evaluation by listeners.

There is a considerable number of variations in signal processing that could be applied to the envelope to mimic the SNR used in an IBM, and hence be used as the basis for implementing a BM. In addition, there is no prior knowledge on how to define and apply a threshold to the metric replacing the SNR, to switch the BM from zero (value of the metric below threshold) to unity (value of the metric above threshold).

The magnitude of the threshold for both IBM and BM algorithms becomes of critical importance when speech is "buried" in noise. If the threshold is set too high, too many subbands will be assigned a gain of zero and the intelligibility could be reduced. Conversely, if the threshold is set too low, too many subbands will be assigned a gain of unity and the output is likely to contain excessive noise. The latter condition may also reduce intelligibility. Of these two potential outcomes, it has been suggested that there is less influence on the speech intelligibility of an IBM, and hence we infer on a BM, if the metric tends to be too high rather than too low (Li & Loizou, 2008).

Examples of the waveforms of speech, and speech in noise, in our first implementation of an IBM and BM are shown in Figure 3 (see next page). Time histories are shown for signals at the output of the algorithms in the upper part of the diagram and for a selected subband below. They extend for 30 seconds and contain seven utterances of the form "Circle the . . [insert test word] . . . again". The test words are different in each utterance. All waveforms in Figure 3 are time aligned (i.e., vertically aligned), so the times when speech occurs can be deduced.

The seven utterances are shown in the absence of environmental noise in the top waveform of the upper part of the diagram (labeled "noise-free speech"). The words repeated at the beginning and end of an utterance can be seen to produce very similar but not always identical waveforms, presumably reflecting slightly different pronunciation or intonation by the speaker.

The overall speech SNR has been chosen so that the speech is completely "buried" in noise, as can be seen by the second waveform in the upper part of the diagram labeled "unprocessed speech in noise". The outputs of two algorithms implementing either an IBM or a BM are shown directly below the unprocessed speech in noise. It is immediately evident from the time periods when there is no speech that the environmental noise is not completely attenuated by the binary masks. This is because initial listening to the outputs of the algorithms revealed unwanted sounds - clicks, musical noise and other distortions - that rendered the quality of the processed sounds unacceptable. For this reason, the "off" value of the masks was set to 0.5 rather than zero, and in this way the residual environmental noise is used to mask the sounds affecting speech quality. Close inspection of the outputs of the two algorithms reveals that some features of the original speech have been recovered from the unprocessed speech in noise, with more features recovered by the IBM than the BM, as would be expected. The ability of a binary mask to recover the original speech depends on the overall speech SNR, which is low in Figure 3 (i.e., see the unprocessed speech in noise in the upper part of the Figure). More features are recoverable when the SNR is greater than that shown in Figure 3, and less when the SNR is less than that shown in the diagram (i.e., more noise, less speech). It should be noted that individual subbands may possess SNRs greater or less than the overall speech SNR, depending on the frequency spectrum of the noise.

Details of the subband signal processing by the binary masks are shown in the lower part of Figure 3. The time histories are examples for a subband in which speech sounds are more intense than the environmental noise. The metrics used to construct the binary masks are: for the IBM, the short-term SNR in the subband (labeled "SNR (dB)" in the lower part of Figure 3); and for the BM, the ratio of envelopes representing an estimate of the speech to an estimate of the combined speech in noise (labeled "magnitude ratio" in Figure 3). The calculation of the magnitude ratio had to be limited to components at low frequencies to achieve a stable metric.

For the IBM, inspection of the waveform for the SNR of this subband reveals that the mask metric produces peaks at all times speech is present in the unprocessed speech in noise (i.e., compare "IBM" with "noise-free speech"). The threshold for activating the ideal mask was set to -5 dB and is shown by a horizontal line in Figure 3. It can be seen to be exceeded, thereby setting the output of the mask to unity, twice during the first, second, fifth and seventh utterances, and once during the third, fourth and sixth utterances. The subband output for the IBM, shown by the waveform below the metric (labeled "subband output after applying IBM"), contains signals with magnitudes greater than the noise that coincide both with the timing of the mask being unity and the occurrence of utterances, and so can be expected to contain speech.

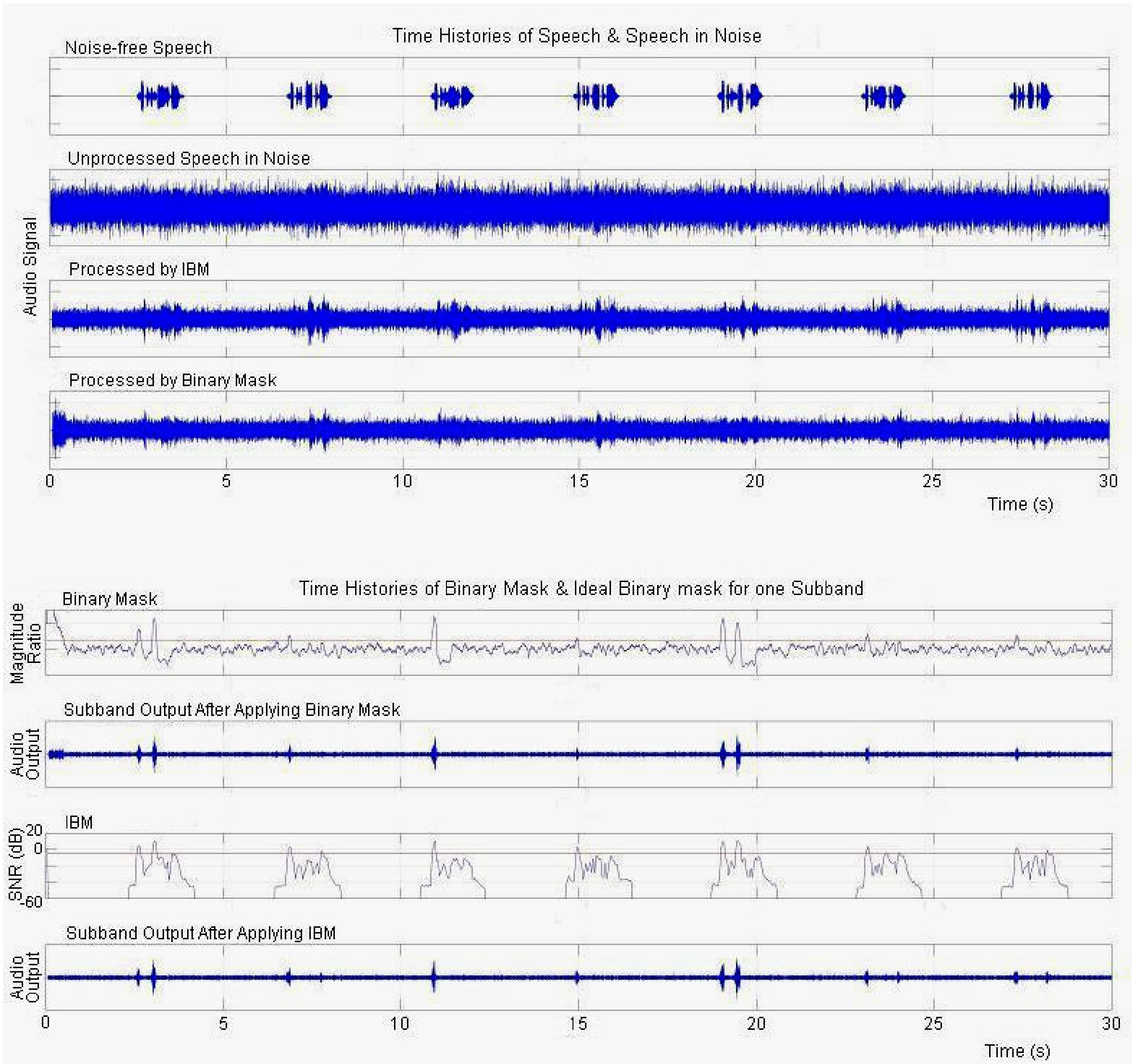


Figure 3: Time-aligned time histories of speech and speech in noise, both unprocessed and processed by either an ideal binary mask (IBM) or a binary mask (BM). Upper - Speech, unprocessed speech in noise, and output of algorithms; Lower - BM and IBM metrics, and outputs for one subband.

Turning to the BM, it is evident from the waveform that this mask metric also produces peaks coincident in time with the utterances (see "binary mask" in the lower part of Figure 3). The threshold for activating the BM is again shown by a horizontal line (at a magnitude ratio of about 1.3). The threshold is selected by trial and error so that sounds dominated by speech exceed the value while sounds dominated by noise do not. In the case shown, where sounds dominated by speech clearly exceed the thresholds of the two masks, not all speech identified by the IBM is identified by the BM. Thus, when comparing the subband output after processing by the IBM or BM, it can be seen that while the large magnitudes are outputted by both masks, some of the smaller features of the audio signal are not detected by the BM (e.g., see the second, sixth and seventh utterances of the "subband output after applying binary mask"). It would therefore appear that the performance of this BM will be inferior to that of the IBM, though the influence on speech understanding cannot be predicted from these data.

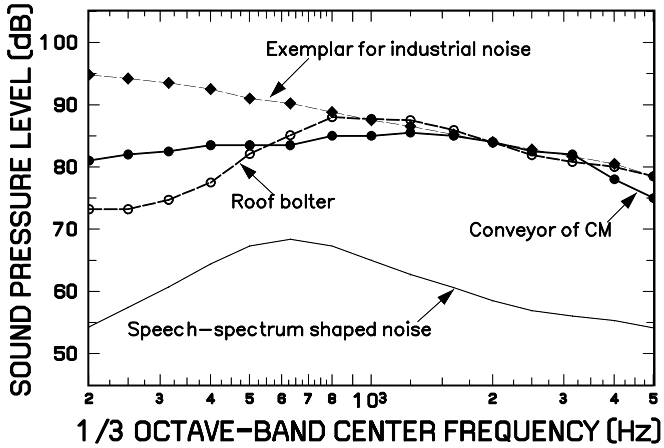


Figure 4: Typical one-third octave band frequency spectra of roof bolter and continuous miner (CM) noise at the operator's ear or near the machine. The spectrum of speech-spectrum shaped noise and an industrial-like noise are shown for comparison.

4.0 Technology Evaluation

4.1 Selection and Simulation of Noises for Listening Tests

Two candidate mining machines have been selected for use in the study. Typical one-third octave-band frequency spectra of the sound pressures experienced by operators or persons nearby the conveyor of a continuous miner and a roof bolter are shown in Figure 4 (Camargo et al., 2016; Szary et al., 2011). In a survey of twelve underground coal mines, the standard deviations of the sound levels produced by 33 continuous miners when cutting and loading, and 37 roof bolters when drilling, were ±2.6 and ±3.2 dBA, respectively (Bobick and Giardino, c1976). Accordingly, the noise spectra in Figure 4 are considered representative of the machine type, and have served as the basis for recordings of sounds for use in listening tests.

Also shown in the figure are one-third octave band frequency spectra for speech-spectrum shaped noise (ECMA TR/105, 2012), and a generalized industrial-like noise. The former is provided for comparison with the spectra of the noises. A time history for the latter was available from another study, and was used in the development and evaluation of algorithms.

The process for simulating time histories of the mine machine noises for presentation to listeners from their frequency spectra starts with a recording of random noise. There are many candidate sources of random noise available on the internet, but only one has been found that meets our requirements for frequency content and purity.

Table 1: Deviations from True Pink Noise

Center Frequency (Hz)	Deviation from pink noise (dB)		Center Frequency (Hz)	Deviation from pink noise (dB)
160	+1		1000	0
200	0		1250	+1
250	0		1600	+1
315	0		2000	+1
400	0		2500	+2
500	0		3150	+2
630	0		4000	+3
800	0		5000	+3

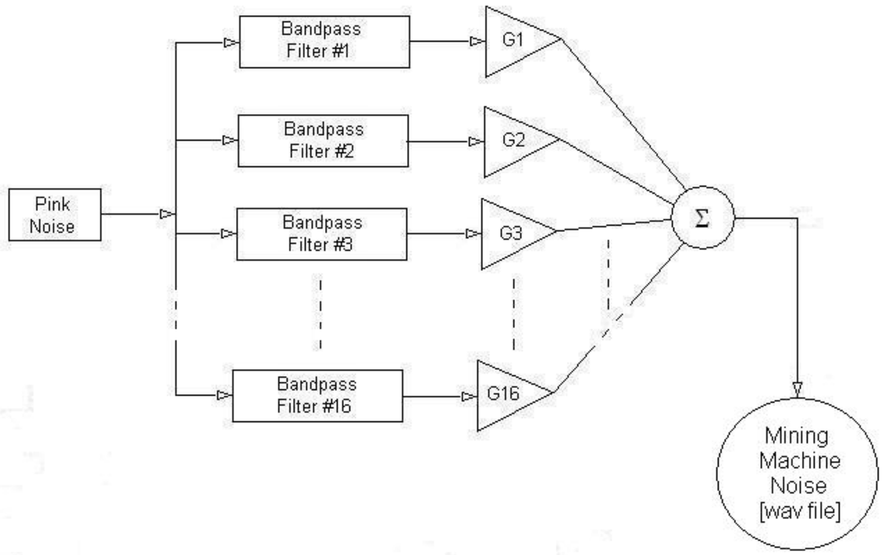


Figure 5: Block diagram showing process for generating simulated mining machine noises from one-third octave band frequency spectra (for details see text).

With the digitized random noise recording selected, attention turned to the construction of a set of sixteen, parallel, one-third octave, band-pass filters, in order to create waveforms with the frequency spectra shown in Figure 4. The filters were constructed in MATLAB as finite impulse response (FIR) digital filters using the concept block diagram shown in Figure 5. The center frequencies are those specified by the American National Standards Institute (ANSI S1.1-2004). As the sampling frequency in all our digital signal processing is 12 kHz, the highest frequency that can be reproduced is 6 kHz. The center frequency of the highest-frequency, one-third octave band that can hence be employed is 5 kHz, which has an upper frequency limit of 5.6 kHz. The lowest frequency band constructed has an upper frequency limit of 178 Hz, which is less than the lowest frequency of the speech sounds used in this study (200 Hz). Thus, our simulations of time histories of mining noises have a usable bandwidth from below the minimum frequency of the signal processing to 5.6 kHz.

The responses of the one-third octave band filters to the pink noise source are shown in Table 1 (see previous page). The Table indicates the deviations from true pink noise introduced by the source and filters combined (i.e., deviations from pink noise would be zero in all bands for *ideal* pink noise and *ideal* one-third octave band filters). Reference to Table 1 shows that there is a small deviation from a "flat" response in the one-third octave band with center frequency at 160 Hz, and increasing deviations in one-third octave bands with center frequencies of 1.25 kHz and above. Perhaps up to 2 dB of the deviation at frequencies of ~3 kHz, and above, can be attributed to imperfections in the sound source. The remaining deviations would appear to originate within the filters. Irrespective of origin, the deviations from flat response in Table 1 serve as the correction factors to apply when producing recordings to simulate the noise of mine machines.

The time histories for the different mine machines can now be obtained by multiplying the output of each one-third octave band filter by the corresponding one-third octave-band value for a machine. The latter are computed from the data in Figure 4 with, additionally, corrections for the deviations from unity of the source and one-third octave band filter set listed in Table 1. This is done by introducing filter-specific gains G1 - G16, as shown in Figure 5. The resulting one-third octave-band spectra to be applied to our pink noise to obtain recordings simulating the time histories of the noise of the machines are given in Table 2. For implementation, the relative gain values are expressed as voltages in dB re a convenient reference, here chosen to be 1 microvolt rms. The adjusted band levels are finally summed to give the full bandwidth simulation of the noise of a given machine (shown by the "Σ" in Figure 5).

Table 2: One-Third Octave-Band Frequency Spectra for Simulating the Noise of Selected Mining Machines

Center Frequency (Hz)	Continuous Miner (Conveyer Starts) (dB re 10^{-6} V rms)	Roofbolter (dB re 10^{-6} V rms)
160	79	72.1
200	81	73.2
250	82	73.2
315	82.5	74.7
400	83.5	77.5
500	83.5	82.1
630	83.5	85.1
800	85	88
1000	85	87.7
1250	84.5	86.5
1600	84	84.9
2000	83	82.9
2500	80.5	79.9
3150	80	78.8
4000	75	77
5000	72	75.5

4.2 Simulation of Sounds Experienced by Miner Wearing a Hearing Protector

A simulation of the sounds that would be experienced by a miner wearing an HPD has been developed. A physical representation of the simulation is shown in Figure 6. Here a worker is wearing a commercial, passive, circumaural HPD that consists essentially of two cups (one per ear) held in place by a spring, with soft cushions sealing the air space between the ear cup and the head. There is also a microphone mounted on the outside of each ear cup. The purpose of the microphone is to sense sounds in the environment around the worker.

Our simulation of the device in the photograph is purely computational: physically, there is no HPD nor microphone. The photograph, however, serves to illustrate that sounds can reach the ear essentially by two paths: 1) by passing through the HPD, or 2) by replaying the output of the microphone through a miniature loudspeaker under the ear cup after signal processing.

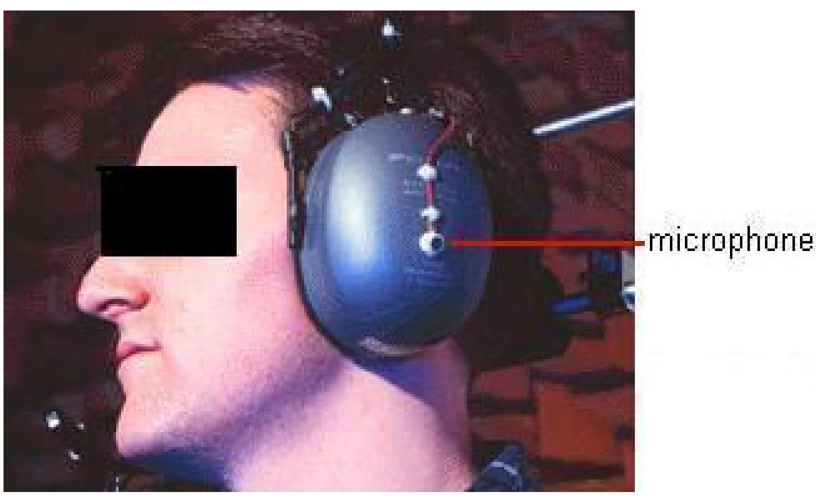


Figure 6: Photograph of worker wearing a passive, circumaural HPD.
A miniature microphone has been attached to the outside of the ear cup.

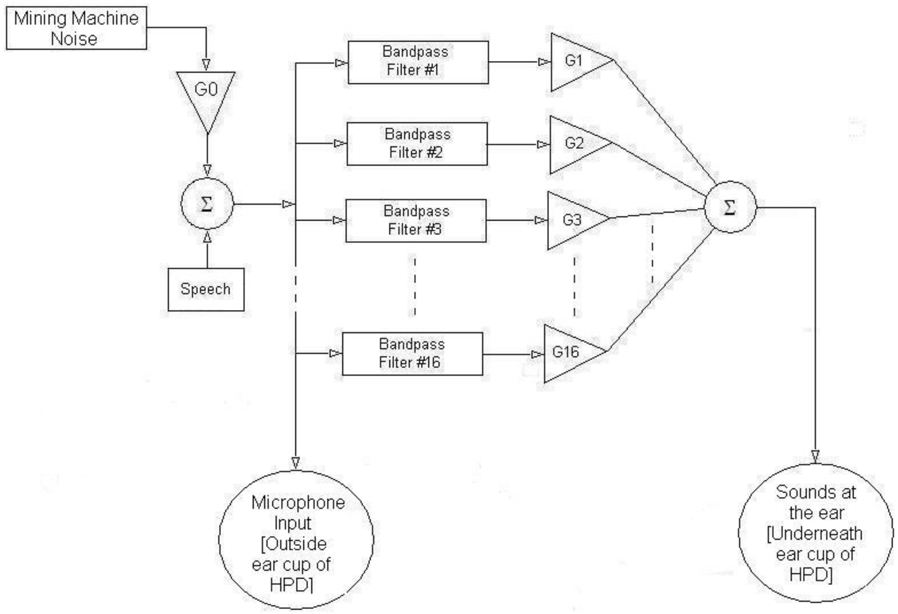


Figure 7: Block diagram showing process for simulating a passive HPD (for details, see text).

The sounds experienced by a miner wearing an HPD are simulated by the process shown in Figure 7. Our model for a conventional, passive HPD consists of a set of sixteen, one-third octave, band-pass filters each followed by an amplifier.

In order to simulate the sounds heard by someone standing close to an operating mine machine while listening to someone else talking, the input to the "hearing protector" consists of speech to which is added the noise of the mine machine (shown to the left of the diagram). The time histories of the noises are obtained for selected machines by the process described in subsection 4.1. The speech consists of sentences from a speech-in-noise test (House et al., 1965), which is described in the following subsection. An amplifier, labeled G0, enables the magnitude of the noise to be changed to set a desired speech SNR, which models how far the talker is from the listener (i.e., the greater the distance the less speech intensity relative to the environmental noise, hence lower SNR).

The attenuation of the HPD to apply to the speech buried in noise is obtained by multiplying the output of each one-third octave-band filter by the corresponding one-third octave-band attenuation of the HPD. The latter are computed from the data in Table 3 (see next page), which were measured previously in our laboratory when subjects wore a high-performance, commercial, passive circumaural HPD. This is done by introducing filter-specific gains G1 - G16 as shown in Figure 7. For implementation, the values are expressed in Table 3 as the attenuation in dB.

The adjusted one-third octave-band levels are finally summed to give the full bandwidth simulation of the sounds at the ear under the ear cup of the HPD (shown by the "Σ" in Figure 7). These are the sounds heard by the miner in the simulation when wearing a commercial circumaural HPD that provides excellent attenuation, such as the model shown in Figure 6. The recordings simulating sounds at the ear under the passive HPD serve as the reference against which the performance of our algorithms is to be assessed in the noise of mine machines.

The sounds to be used as input to our algorithms consist of the speech and noise in the environment at the location of the listener (i.e., close to a mine machine), and would be obtained in the real world from the microphone outside the ear cup. In our simulation, these are before the filter set and are easily recovered after the mine machine noise and speech have been combined (see Figure 7). To simulate a remote talker, speech and noise are inputted separately to the algorithm (see also Figure 1).

Table 3: One-third Octave Band Attenuation of a Commercial, Circumaural HPD

Center Frequency (Hz)	Attenuation of HPD (dB)		Center Frequency (Hz)	Attenuation of HPD (dB)
160	13.5		1000	32.5
200	17		1250	35.5
250	21.5		1600	36
315	24		2000	35.5
400	27.5		2500	36.5
500	30		3150	35
630	31		4000	40.5
800	32		5000	43

4.3 Subjects for Listening Tests

Healthy volunteers were invited to participate in the study after undergoing an induction procedure. Criteria for inclusion in the study were: 1) age between 18 and 50 years; 2) absence of factors that could influence the performance or acceptability (e.g., comfort) of insert earphones, or circumaural headphones; 3) no infections of the skin or external ear; 4) no impacted cerumen; 5) no middle ear infections; 6) tolerance to plastics on the skin; 7) hearing function questionnaire score less than 10 (see Appendix *Hearing Function Questionnaire*); 8) when measured, pure-tone air-conduction Hearing Threshold Levels (HLs) re ANSI S3.6-1996 of less than (i.e., more sensitive than) 20 dB HL at frequencies of 0.25, 0.5, 1, 2, 4, 6 and 8 kHz, with thresholds of individual ears that differ by less than 10 dB (Schlauch and Nelson, 2009); 9) absence of persistent tinnitus, and; 10) ability to identify similar sounding words spoken out of context in American English, read the words, and signal the words selected by completing a written form. Volunteers successfully completing the induction procedure were accepted as subjects for as many listening tests as they chose and were paid for their time.

Audiometric evaluation of hearing thresholds was conducted on volunteers who attended our audiology clinic, as well as an examination of their external ears. For other subjects, information concerning their hearing ability was obtained from answers to the Hearing Function Questionnaire.

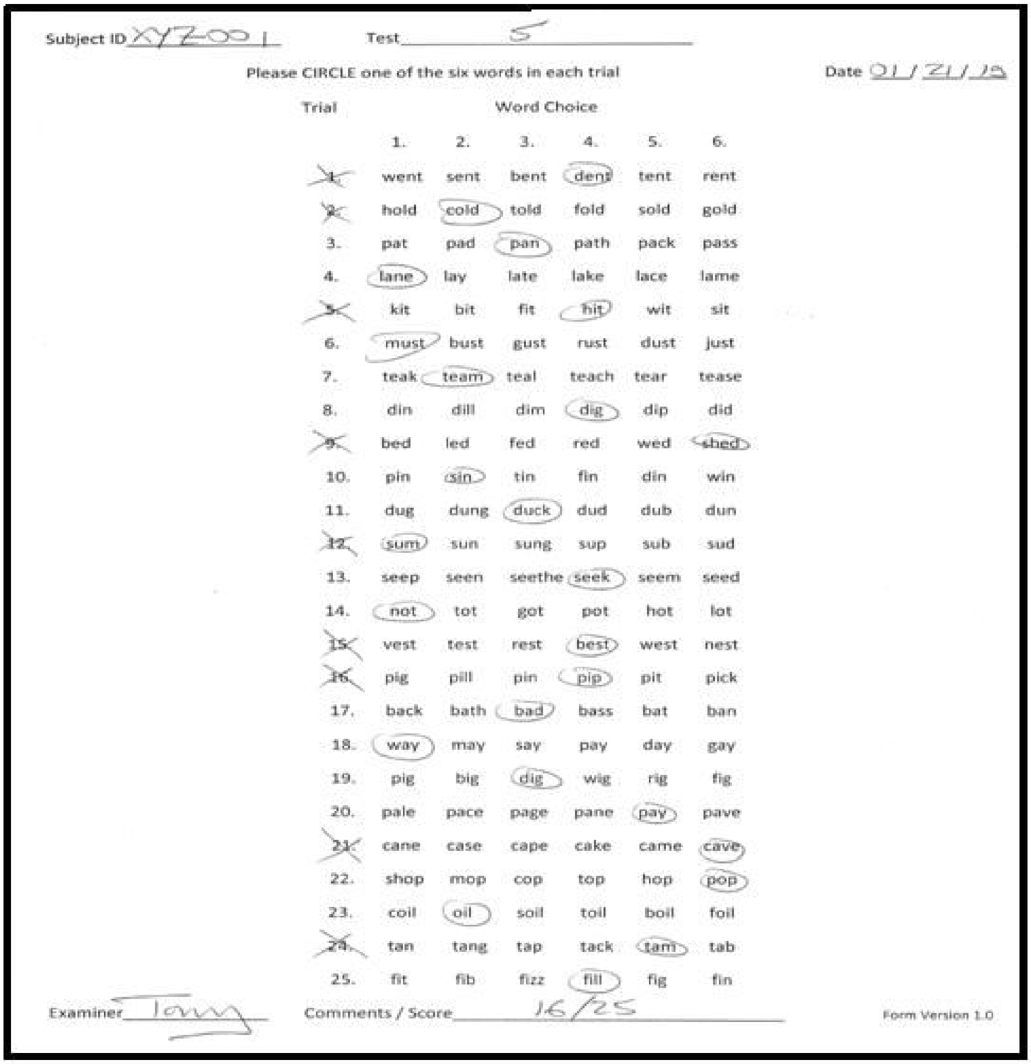
Subjects who came to our clinic and/or laboratory at the University of Connecticut Health Center were seated comfortably in an audiometric room, which fulfilled the ambient noise requirements for audiometric threshold determinations in ANSI S3.1, 1999. The listening tests were conducted with the subject wearing commercial insert earphones (E·A·RTone type 3A or 5A) using .wav files constructed off-line by MATLAB. Subjects who listened remotely to the .wav files used their choice of headphones or earphones. The influence of listening conditions on the performance of listening tests is discussed in the following subsection.

The study protocol was approved by the Institutional Review Board of the University of Connecticut Health Center, Farmington, CT.

4.4 Listening Tests - On Campus, and Using the Internet

As in our previous work for the Alpha Foundation, the Modified Rhyme Test (MRT) was used to characterize the intelligibility of individual words in a six-alternative forced choice paradigm (ANSI, 2009; House et al., 1965). This test of consonant confusion has been used extensively for its relevance to critical communications in which a single word error could have serious consequences (e.g., air traffic control, military and first responder operations) (Cardosi, 1998; Anderson et al., 1997; LaTourette et al., 2003), and so is most suitable in our opinion for simulating situations that may occur in mining environments. It was also used in the only evaluation of eHPDs for miners (Azman and Hudak, 2011).

The word lists were those standardized for American English as spoken by a male talker (Auditec, St Louis). A trial consisted of one of six words being randomly replayed within a carrier sentence, e.g., "Circle the - [insert test word] - again". Subjects were instructed to



Subject ID XYZ001 Test 5

Please CIRCLE one of the six words in each trial

Date 01/21/19

Trial	1.	2.	3.	4.	5.	6.
~~1.~~	went	sent	bent	dent	tent	rent
~~2.~~	hold	cold	told	fold	sold	gold
3.	pat	pad	pan	path	pack	pass
4.	lane	lay	late	lake	lace	lame
~~5.~~	kit	bit	fit	hit	wit	sit
6.	must	bust	gust	rust	dust	just
7.	teak	team	teal	teach	tear	tease
8.	din	dill	dim	dig	dip	did
~~9.~~	bed	led	fed	red	wed	shed
10.	pin	sin	tin	fin	din	win
11.	dug	dung	duck	dud	dub	dun
~~12.~~	sum	sun	sung	sup	sub	sud
13.	seep	seen	seethe	seek	seem	seed
14.	not	tot	got	pot	hot	lot
~~15.~~	vest	test	rest	best	west	nest
~~16.~~	pig	pill	pin	pip	pit	pick
17.	back	bath	bad	bass	bat	ban
18.	way	may	say	pay	day	gay
19.	pig	big	dig	wig	rig	fig
20.	pale	pace	page	pane	pay	pave
~~21.~~	cane	case	cape	cake	came	cave
22.	shop	mop	cop	top	hop	pop
23.	coil	oil	soil	toil	boil	foil
~~24.~~	tan	tang	tap	tack	tam	tab
25.	fit	fib	fizz	fill	fig	fin

Examiner Tony Comments / Score 16/25 Form Version 1.0

Figure 8: Example of a completed 25-trial Modified Rhyme test.

identify the test word on a prepared word list. There were 25 trials in each test from which a word score (i.e., number of words correctly identified) was derived for a preset speech SNR. An example of a completed test is shown in Figure 8.

A successful demonstration of proof-of-concept would be obtained by an increase in word score when an algorithm is employed compared to a reference condition when it is not employed, thus demonstrating improved speech intelligibility. The statistical test of the difference between the word scores for the two conditions is a two-sided paired t-test (Bland, 2015).

As we previously reported, on-campus listening tests require subjects to come to our clinic and/or laboratory at the University of Connecticut Health Center. Unfortunately, testing was disrupted by the COVID-19 pandemic. It became apparent from the response to advertisements that volunteers were commonly unwilling to come on campus to undergo listening tests, especially during the time when employees and students were encouraged to work from home. For this reason, we developed a web-based listening test employing the MRT. The scientific challenges to web-based speech intelligibility tests involve the following issues.

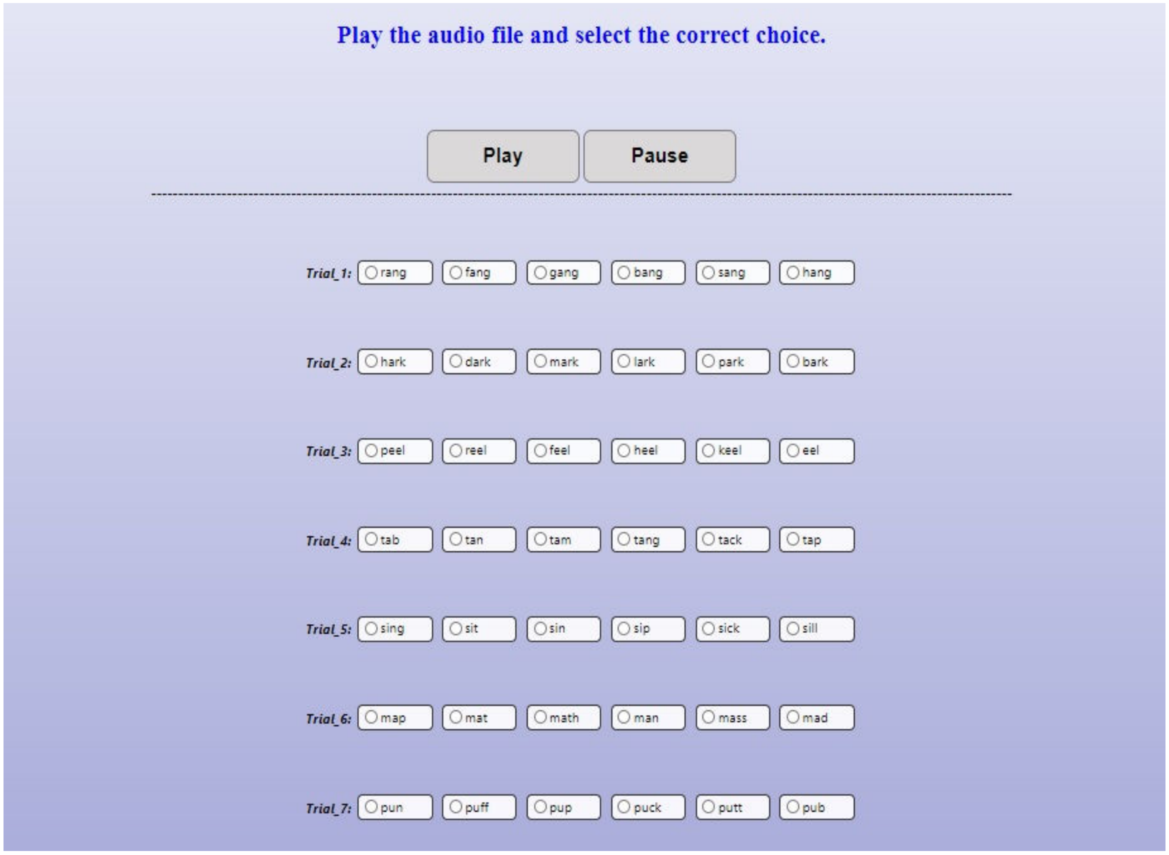


Figure 9: Computer screen for test webpage of internet-based MRT showing trials #1 - #7.

Firstly, do listeners who may never visit our campus possess normal hearing? We have attempted to address this issue by restricting the age of subjects, and by including the short questionnaire on hearing ability (see Appendix *Hearing Function Questionnaire*).

Secondly, does the fidelity of the sound reproduction system chosen by the listener influence the performance of the test? Clearly, the answer to this question is yes, but the important consideration is the extent to which any detrimental effects can be reduced. Commercial headphones, earphones and earbuds generally possess acceptable responses at the frequencies most important for understanding speech (i.e., from about 500 Hz to 2-3 kHz). Note that this statement refers to speech intelligibility but not speech quality, which will be influenced by the restricted frequency range. The computer loudspeaker is not used. Moreover, defining the proof-of-concept demonstration as the *difference* between two tests performed during the same session, that is one without our algorithm and one with, may be expected to reduce effects due to spectral imperfections in the sound reproduction system.

Thirdly, will the lack of control of the sound level listeners choose for their listening test influence the results? It is well known that speech intelligibility is not influenced by the loudness of sounds over a wide range of sound levels (Dubno et al., 2005), so provided listeners choose a "comfortable" listening level (i.e., not too loud or too quiet) and can hear all sounds the effect on word scores is unlikely to be detectable. For this reason, our instructions to users include examples of the quietest and loudest sounds for them to set the volume control of their sound reproduction system before commencing formal listening tests.

The screen seen by a subject on commencing an internet-based listening test is shown in Figure 9.

Subjects started the test when they chose by mouse-clicking on the "play" button and may temporarily stop the test at any time during the 25 trials by mouse-clicking on the "pause" button. Once the test started, MRT sentences were replayed once for each trial, with a ~3 s interval between each trial for subjects to record their choice by mouse-clicking on the word they believed they heard. Subjects were asked in the instructions to respond to each trial even if they had to guess the word when the speech was muffled or unclear. They also had to scroll down during the test to answer trials that were not initially visible on the screen (e.g., trials #8 *et sequa* for the webpage shown in Figure 9). After completing all the tests in the measurement session, subjects were requested to e-mail the response file to us for analysis. They were sent payment for their time on receipt of the data file.

4.5 Analysis of Listening Tests

We have during the course of the project conducted listening tests under three different measurement conditions: in an audiometric clinic, a research laboratory, and remotely using the internet.

The first, described as clinic listening tests, were conducted under controlled conditions within a controlled environment in the audiometric facilities of the Division of Otolaryngology, Head and Neck Surgery. The clinic is operated by experienced audiologists, who examined the external ears of subjects, supervised the fitting of insert earphones and controlled the sound level of stimuli prior to presentation. The hearing thresholds of each subject were first established at different frequencies, and sounds were then presented at a fixed intensity relative to an individual's thresholds. Measurements were thus conducted at the same *sensation level* for all subjects. Subjects recorded the responses to each trial on paper (e.g., see Figure 8), including guesses when words were unclear or inaudible and the test proceeded at a rate established by agreement between the experimenter and subject.

The second, described as laboratory listening tests, were conducted in an audiometric room, in our laboratory, which also provided a controlled environment. Subjects themselves inserted earphones into their ears, with some oversight from graduate students. There was no evaluation of hearing thresholds or use of constant sensation level. As in the clinic, subjects recorded the responses to each trial on paper, including guesses when words were unclear or inaudible, and the test proceeded at a rate established by agreement between the experimenter and subject.

The third, described as web-based listening tests, were conducted at a location of the subjects' choice, commonly at home, and they chose the headphones or earphones to wear. They listened to sounds reproduced by their own computer's audio and accessed sounds by mouse-clicking on symbols visible on their computer screens. At all times subjects controlled when the next listening test would start and could temporarily stop the test at any time by mouse-clicking on a "pause" button (see Figure 9). Once a test started, MRT sentences were replayed once for each of the 25 trials and subjects were asked to respond to each trial even if they had to guess the word when the speech was unclear or inaudible, as in the other measurement conditions.

It is not evident *a priori* that the three measurement conditions will produce equivalent results. Of the three, the most controlled conditions were those employed in the clinic, henceforth taken as the "gold standard" with which the results of other listening tests will be compared.

Before comparing mean word scores between different measurement conditions, it is important to identify whether the comparative lack of control of subject performance during laboratory or web-based testing (e.g. poor fitting earphones or headphones, imperfect sound reproduction, lack of concentration, interruptions, etc.) could lead to systematic differences in subjects' word scores.

The identification of systematic, as opposed to random, errors is an imperfect art. However, by conducting listening tests consisting of the same sequence of audio files, each containing the same 25 trials, it was possible to separate random differences in word scores between subjects from the systematic, or erratic, performance of a subject during a series of tests. An example of the method employed is shown in Table 4 (next page).

Table 4 Subject's Word Scores in an Experiment Conducted in the Clinic and Laboratory Consisting of 13 Tests

Test Location - Clinic														
	Test Number(#)													Summed
Subject #	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	Scores
1	25	23	19	17	24	23	19	25	23	20	24	24	20	286
2	24	22	19	15	22	23	16	24	23	18	24	25	20	275
3	25	20	19	15	22	22	18	21	21	17	21	23	17	261
4	25	21	19	18	21	20	17	25	21	20	23	23	21	274
5	25	24	21	13	23	23	15	24	24	17	24	24	18	275
6	25	22	19	14	24	21	16	23	21	19	24	22	20	270
7	24	23	20	14	23	22	18	22	23	19	25	23	19	275
Mean Summed Score All Tests														274
SD														7.4
Score (%)	98.9	88.6	77.7	60.6	90.9	88.0	68.0	93.7	89.1	74.3	94.3	93.7	77.1	
SD	2.0	5.4	3.1	7.1	4.5	4.6	5.7	6.0	5.0	5.1	5.1	3.9	5.5	
Test Location - Laboratory														
8	25	21	18	14	22	20	19	23	23	17	23	23	21	269
9	25	23	21	13	21	23	17	19	23	20	21	22	18	266
10	25	21	21	18	21	21	18	20	20	20	23	22	19	269
11*	**25**	**17**	**17**	**14**	**22**	**18**	**14**	**19**	**20**	**10**	**23**	**22**	**12**	**233**
12	25	23	20	19	23	21	16	23	23	17	23	21	18	272
13	25	23	20	18	25	24	19	23	21	17	23	23	19	280
Mean Summed Score All Tests														265
SD														16.3
Score (%)	100.0	85.3	78.0	64.0	89.3	84.7	68.7	84.7	86.7	67.3	90.7	88.7	71.3	
SD (%)	0.0	9.4	6.6	10.4	6.0	8.5	7.8	8.2	6.0	14.6	3.3	3.0	12.2	
Mean Summed Score Without Outlier														271
SD														5.4
Score (%)	100.0	88.8	80.0	65.6	89.6	87.2	71.2	86.4	88.0	72.8	90.4	88.8	76.0	
SD (%)	0.0	4.4	4.9	10.0	8.2	7.3	3.6	7.8	5.4	6.6	3.6	2.2	4.4	

A complete record of subjects' word scores, either in the clinic or the laboratory, listening to speech in continuous miner or roof bolter noise is shown in Table 4. In this experiment there were seven subjects who attended the audiology clinic and six who attended the laboratory. Subjects listened to the same audio files either in the clinic or the laboratory. There were thirteen different tests that were administered sequentially in the same order to each subject during a measurement session. In Table 4 the tests are numbered from #1 - #13 with each involving a different mining noise or speech SNR except for test #1 which consisted of speech with no noise. This test was to familiarize the subject with the measurement procedure: a perfect score of 25 words correct was expected for each subject. Reference to Table 4 shows that all subjects attending the laboratory achieved a perfect word score. However, two subjects attending the clinic only achieved a word score of 24 / 25 (subjects #2 and #7). The difference in mean word scores is not statistically significant although it may reflect a transitory lack of concentration on the part of the two subjects with lower scores.

If the magnitudes of the standard deviations (SDs) for tests from #2 to #13 are reviewed for all subjects, it can be seen that they range from 3.1 to 7.1 in the clinic but from 3.0 to 14.6 in the laboratory. The greatly increased maximum SD in the laboratory might reflect the reduced control of the measurement, though this seems unlikely given the similarity of the minimum SDs, or indicate the presence of one or more subjects with consistently unusual performance whose word scores form outliers in some, but not necessarily all, test conditions.

The analysis to identify a subject with unusual performance is performed in the following way. We can expect that each test will have a most probable value, which will be described by the mean of a near Gaussian distribution (for mean scores not censored by zero or 25). Thus, individual subject's scores will range randomly with respect to the mean for each test. For some tests, a subject will score higher than the expected value and for others lower. If we now sum the scores *for a subject across all tests* (i.e., sum values by row in the table), the properties of the mean summed scores can be employed to identify outliers. By summing word scores across tests for a given subject, the random deviations from the expected values of individual tests will be effectively reduced in the magnitude of the summed mean score, as the summed deviations from expected values will tend to zero as the number of tests increases. Residual differences between subjects in scores summed across all tests will then indicate differences in a subject's performance.

The scores summed over all tests are given for each subject in the far right column of Table 4. Inspection of these data reveals that the mean summed word score is 274 for all subjects who attended the clinic and 265 for all subjects who attended the laboratory. While this difference could reflect an overall difference in performance for subjects attending the laboratory versus the clinic, the origin of the difference in this case is to be found in the SDs: the SD of the summed scores in the clinic is 7.4, while that for the laboratory is much larger, namely 16.3. Close inspection of the summed scores for subjects attending the laboratory identifies one subject, #11, whose summed score is much less than the mean (233 versus 265). If this subject is removed from the analysis, the mean summed score for the remaining subjects attending the laboratory increases to 271 and the SD decreases from 16.3 to 5.4 (see bottom of far right column of Table 4). Clearly, eliminating subject #11 from the analysis has removed the discrepancy in the SDs of the mean summed scores and reduced the difference between the mean summed scores to about 1%. The summed score of the outlier is hence revealed as being ~7 SDs from the mean summed score after it has been removed from the calculation. The probability of the outlier having obtained meaningful word scores is thus vanishingly small.

The effect of removing the outlier on the results of tests #2 - 13 conducted in the laboratory can be seen by comparing the mean scores and SDs for each test at the bottom of the table. While reducing the number of subjects from six to five would usually be expected to *increase* the SD, a comparison of the SDs including and excluding the outlier reveals that in most cases the SD is reduced when the outlier is excluded: viz., test #2, from 9.4 to 4.4; test #7, from 7.8 to 3.6; test #10, from 14.6 to 6.6; and test #13, from 12.2 to 4.4. More importantly, the largest differences between the mean values for individual tests conducted in the clinic versus the laboratory are reduced when the outlier is excluded: viz., test #10, from 7.0 to 1.5; and test #13, from 5.8 to 1.1.

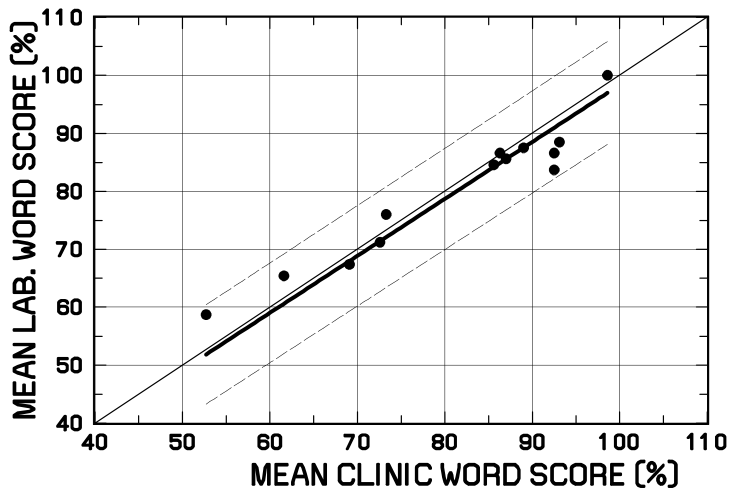


Figure 10: Mean word scores (%) for the same MRT test conducted either in the laboratory or the clinic (filled circles). Thick line is a linear regression fit to the data, and dashed lines are 95% confidence intervals. (for further explanation, see text)

it would now appear appropriate to determine whether listening tests conducted in the laboratory are in agreement with those conducted in the clinic. First, however, it is necessary to account for subjects' responses that involve chance or guessing the correct word during a trial. It is generally recognized that a psychophysical test forcing subjects to choose one item out of several alternatives presented to them, such as the MRT used here, requires subjects to guess the correct word in circumstances in which speech is not clear or inaudible because of the noise. Consequently, it is necessary to adjust the observed word scores for guessing. This has been done to the data of Table 4 using the formula provided for this purpose by ANSI S3.2-1989 (R2009):

$$W_{cor} = W_{obs} - (N - W_{obs}) / (n - 1) \qquad (1)$$

where W_{obs} is the word score recorded by a subject (i.e., number of words correctly identified in a test), W_{cor} is the word score corrected for guessing, N is the number of trials in a listening test (i.e., 25), and n is the number of alternative words from which the subject may choose (i.e., 6).

After correcting the individual word score for guessing and excluding the outlier, the results of the thirteen listening tests in Table 4 are summarized in Figure 10. In this diagram the mean word scores are expressed as the percentage of words correctly identified. Those obtained in the clinic are plotted on the abscissa and those in the laboratory on the ordinate. In this way the mean word scores obtained listening to the same audio file comprising one test of 25 trials can be represented in Figure 10 by a data point (shown as a filled circle). Thus, if the mean word score recorded by subjects listening to an MRT in the laboratory is identical to that recorded by subjects listening to the same test in the clinic, then the filled circle will fall on the thin line bisecting the graph diagonally (from coordinates 40,40 to 110,110). Discrepancies between word scores obtained in the laboratory test compared to the "gold standard" will hence appear as deviations from the thin diagonal line.

Inspection of Figure 10 reveals that word scores in seven of the thirteen independent speech-in-noise tests cluster closely around the thin line, indicating good agreement between the results of laboratory and clinic tests. The others vary in their proximity to the diagonal line and display the magnitude of the deviations that were obtained with the small numbers of subjects. There is, however, generally no statistically significant difference between the results of the laboratory and clinic tests, though the results of test #12 did reach significance ($p<0.05$, two-sided t-test). This occurred because of the uncommonly small SDs recorded in this test (4.7% and 2.6% in the clinic and laboratory tests, respectively).

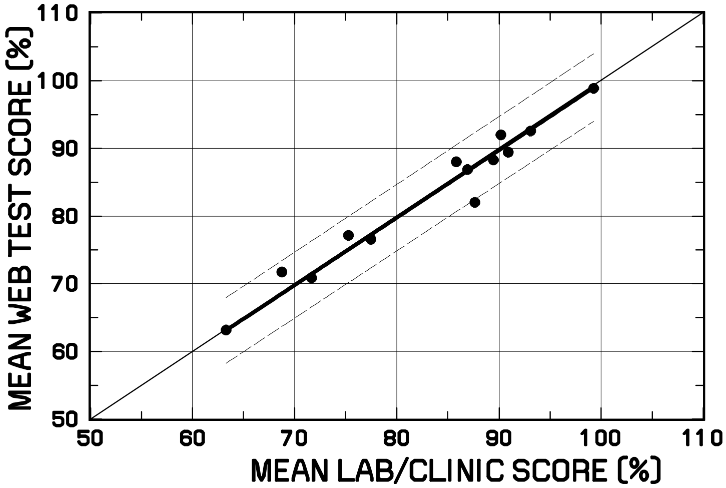


Figure 11: Mean word scores (%) for the same MRT test conducted either on campus (laboratory and clinic) or remotely using the internet (filled circles). Thick line is linear regression fit to data, and dashed lines are 95% confidence intervals. (for explanation, see text)

A linear regression analysis, shown by the thick line in Figure 10, closely follows the desired relation (i.e., the thin line), demonstrating that there appears to be a functional relation between the results of the laboratory and clinic tests. The 95% confidence intervals for the relation are shown by the dashed lines and include all data as well as the desired relation between the laboratory and clinic tests.

Having demonstrated that measurements performed in the clinic and laboratory produce equivalent word scores in more than 90% of tests provided the data are controlled for outliers, it is now appropriate to determine whether listening tests conducted in the clinic and laboratory are in agreement with those conducted remotely using the web-based testing methodology. This has been done in the same way as that just described for comparing word scores obtained in the clinic with those in the laboratory. For this comparison of testing methodologies, there were eleven subjects who underwent the MRT on campus and fourteen who participated remotely. There were thirteen listening tests in all.

The results are shown in Figure 11 where, as before, the mean word scores are expressed as the percentage of words correctly identified. Those obtained on campus are plotted on the abscissa and those off campus on the ordinate. In this way the mean word scores obtained listening to the same audio file comprising one test of 25 trials is represented in Figure 11 by a data point (shown as a filled circle). Discrepancies between word scores obtained by subjects using the web-based test compared to those obtained in the conventional way on campus will hence appear as deviations from the thin diagonal line (from coordinates 50,50 to 110,110).

Inspection of Figure 11 reveals that word scores in twelve of the thirteen independent speech in noise tests cluster closely around the thin line, indicating good agreement between the results of web-based and laboratory / clinic tests. The excellent performance of the web-based listening test for persons with normal hearing is confirmed by a linear regression analysis, shown by the thick line in Figure 11. This closely follows the desired relation (i.e., the thin line), demonstrating there is a functional relation between the results of web-based and laboratory / clinic tests. The 95% confidence intervals for the word scores are shown by the dashed lines, and include all data as well as the desired relation between web-based and laboratory / clinic tests. The maximum deviation in this validation of the performance of the web-based listening test occurred in one test that reached the 95% confidence interval. However, there was no statistically significant difference between the results of web-based and laboratory / clinic tests (two-sided t-test).

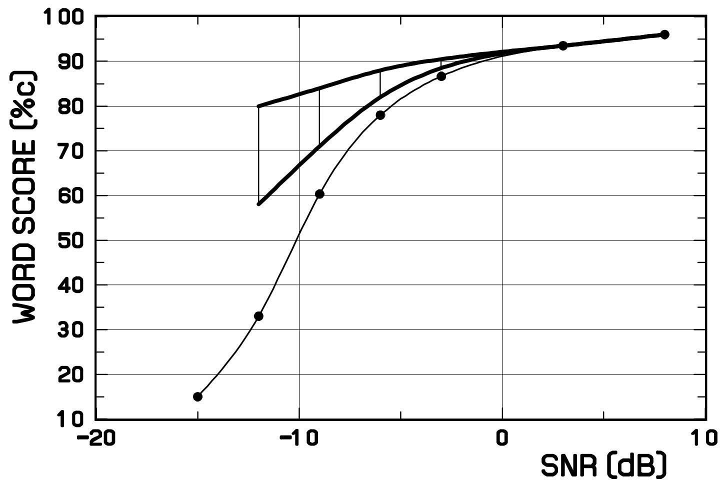


Figure 12: Mean word scores (% correct) shown as a function of the speech SNR. Word scores in the absence of signal processing are illustrated by the filled circles and thin line. Target range for word scores after processing designed to improve speech intelligibility is suggested by thick lines and vertical shading.

In summary, it appears that the web-based test is an adequate alternative to performing tests in the laboratory or clinic. The present work confirms the validity of pooling the results of listening tests conducted on and off campus for persons with normal hearing. With many persons hesitant to come on campus during the COVID-19 pandemic, web-based testing provided a modality for the continuation of listening tests and has been essential for us to address our mission statement. Nevertheless, the necessity to evaluate substantive changes to algorithms by listening tests together with delays, often of many days, before subjects chose to perform a web-based test have inevitably restricted our ability to make progress. It should be noted that listening tests for each experiment took about a month to complete.

4.6 Evaluation of Algorithms for Improving Speech Intelligibility

The primary task of this study was to complete the proof-of-concept by refining algorithms and demonstrating their performance under the conditions requested by the Alpha Foundation. These are described in the mission statement (subsection 2.1). The proof-of-concept evaluation of algorithms consisted of listening tests conducted using .wav files constructed off-line by MATLAB. These were conducted when the subject wore commercial high-fidelity insert earphones on campus, and headphones or earphones of their own choice off campus. In consequence, the components employed for the proof-of-concept evaluation were *ad hoc* and will not form part of a future working prototype.

Listening tests have been performed when subjects listened to speech in an industrial-like noise without hearing protection, to enable comparison with the results of another study, and in mine machine noise when "wearing" the simulation of a passive HPD shown in Figure 6.

4.6.1 Presentation of Results

The results of listening tests are presented as the mean percentage of words correctly identified by subjects in an experiment (%c). They have been obtained for different speech SNRs and different noises to address the mission statement. The presence of systematic errors in word scores has been tested by the method described in subsection 4.5 and data from subjects failing the test have been excluded. In all experiments word scores were obtained when an algorithm was employed compared to a reference condition when it was not employed. The statistical test of the difference between the word scores for the two conditions was a two-sided paired t-test, from which an improvement in intelligibility is obtained when $p < 0.05$.

An example showing the expected form of the results is shown in Figure 12.

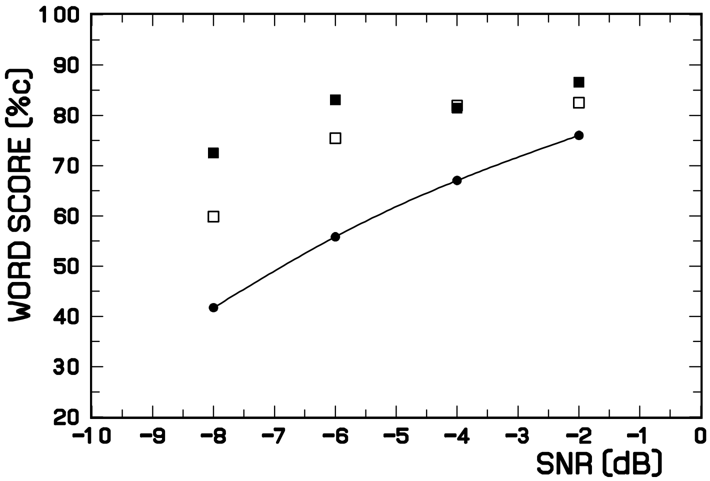


Figure 13: Mean word scores (% correct) at various speech SNRs in the industrial-like noise showing effect of increasing number of subbands. Scores in the absence of signal processing are shown by filled circles and thin line. Word scores after processing using 16 subbands are shown by open squares, and 24 subbands by filled squares.

The word score in the absence of signal processing is illustrated by the filled circles and thin line in Figure 12, and can be seen to follow an 'S'-shaped curve as the SNR increases. The word score is small for large negative values of SNR at which few, if any, words would be understood, improving to almost every word being understood at SNRs of ~10 dB. The target performance of the signal processing performed by our algorithms is suggested by the thick lines and vertical shading in the diagram. It can be seen that the word scores after processing will differ little from those obtained in the absence of signal processing when the latter are greater than about 80%. More importantly, the algorithms should not decrease the word scores under these conditions. Positive contributions to word scores are expected from the signal processing as the SNR decreases (i.e. becomes more negative) until the unprocessed word score is approximately 35%. Under these conditions, an algorithm will need to increase the word score by in excess of 50% to return the word score to more than 80%. Such large increases in word score would appear beyond the reach of algorithms containing up to 24 subbands without access to remote computational resources (e.g., neural network processing, as in Kim et al., 2009). Hence the word scores within the shaded area of the Figure were chosen here as the targets for implementation in a wearable stand-alone eHPD. From a user's perspective, algorithms that produce word scores within the shaded area will provide substantial improvements understanding speech under all conditions of interference by noise.

4.6.2 Comparison of Word Scores for Algorithms Containing 16 and 24 Subbands

It has been argued in subsection 3.2 that algorithms for improving speech intelligibility will be sensitive to the bandwidth of the subbands because of the nature of masking by noise. This may be most readily examined here by constructing an IBM algorithm containing twenty-four subbands in order to compare with the results of a previous study employing sixteen subbands.

Listening tests were undertaken by fifteen subjects who possessed normal hearing in the industrial-like noise to permit comparison with the previous study (mean age 29.6 years, range 21 - 44 years). The results are shown in Figure 13. In this diagram mean word scores (% correct) at various speech SNRs in the industrial-like noise are shown for algorithms containing either sixteen or twenty-four subbands. Mean scores in the absence of signal processing are shown by filled circles and the thin continuous line. Mean word scores after processing using sixteen subbands are shown by open squares, and twenty-four subbands by filled squares.

It is evident by comparing the open and filled squares in Figure 13 that the twenty-four subband algorithm outperformed the sixteen subband version of the algorithm at low SNRs (i.e., more negative) where substantial improvement in speech intelligibility is desired. At -8 dB SNR,

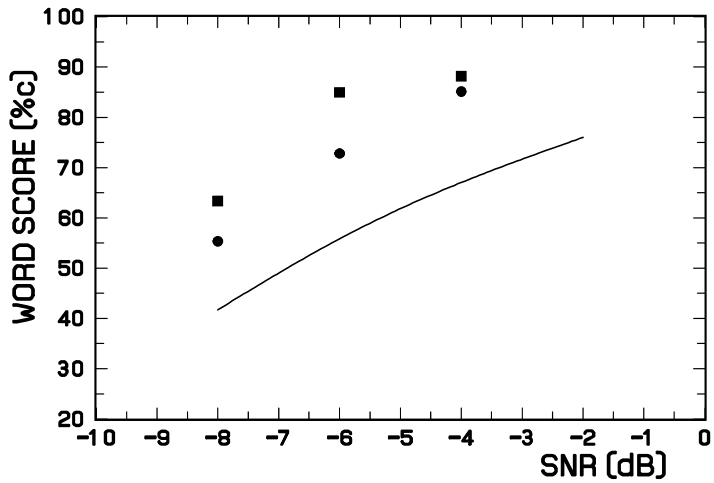


Figure 14: Mean word scores (% correct) at various speech SNRs in continuous miner noise. Scores when wearing a passive HPD are shown by filled circles. Word scores after processing using a 24-subband DM algorithm are shown by filled squares. The thin continuous line shows the word score without signal processing in the industrial-like noise.

for example, the mean increase in word score for the twenty-four subband algorithm was 30.7% while that for the sixteen subband version was 18.1%. Accordingly, all results presented in the remainder of this section were obtained using twenty-four subband algorithms.

4.6.3 Performance of 24-Subband Direct Modulation Algorithm in Mine Machine Noises

Listening tests were undertaken in the simulated noise of a continuous miner or roof bolter by thirty subjects who possessed normal hearing (mean age 26.6 years, range 21 - 42 years). The results are shown in Figures 14 and 15, respectively. In this diagram mean word scores (% correct) at various speech SNRs in mine machine noises are shown as experienced by a miner wearing a passive HPD (filled circles) and when listening to our direct modulation (DM) algorithm (filled squares). The thin continuous line shows the word scores obtained in the industrial-like noise without signal processing.

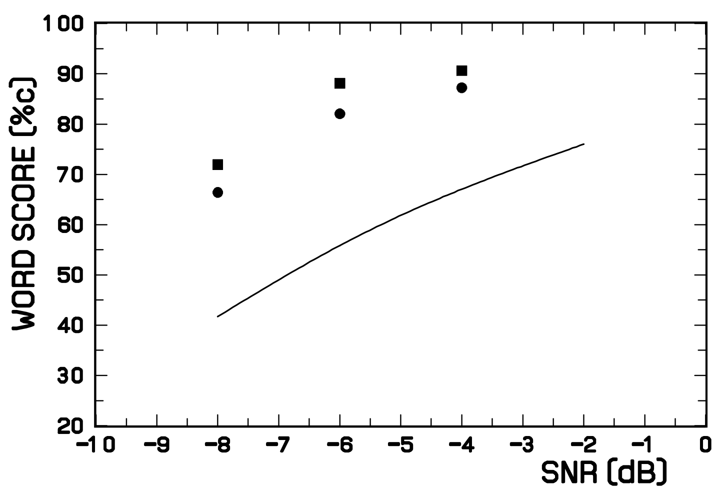


Figure 15: Mean word scores (% correct) at various speech SNRs in roof bolter noise. Scores when wearing a passive HPD are shown by filled circles. Word scores after processing using a 24-subband DM algorithm are shown by filled squares. The thin continuous line shows the word score without signal processing in the industrial-like noise.

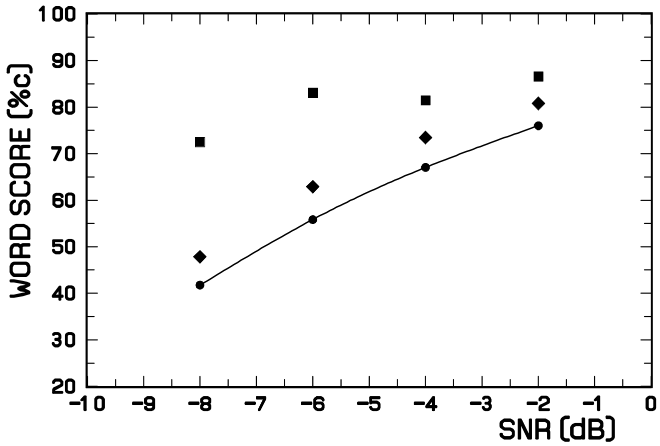


Figure 16: Mean word scores (% correct) at various speech SNRs in the industrial-like noise. Scores in the absence of signal processing are shown by the filled circles and thin line. Word scores after processing using 24 subband algorithms are shown by filled diamonds for a BM, and filled squares for an IBM.

Subjects unexpectedly performed better when wearing the HPD in mine machine noise than previously when listening to the unprocessed industrial-like noise, which led to larger word scores for the comparison test condition. Reference to Figure 12 reveals that only small improvements in word score can be expected with signal processing when the unprocessed word score is >80% (e.g., at an SNR of -4 dB), which is as observed in Figures 14 and 15. Somewhat larger increases in word scores were obtained at lower SNRs with the maximum improvement obtained by the DM algorithm reaching 12.2%. All increases in word scores obtained by employing the algorithm were statistically significant except for the test at an SNR of -4 dB in continuous miner noise when $p = 0.09$.

4.6.4 Performance of 24-Subband Binary Masking Algorithms in the Industrial-like Noise

These listening tests were a continuation of those described in subsection 4.6.2, and involved the same subjects. The purpose of the experiment was to compare the performance of our first ideal binary mask (IBM) with that of our first binary mask (BM).

Mean word scores for subjects listening to speech in the industrial-like noise either unprocessed or processed by our BM or IBM are shown in Figure 16. As before, mean word scores in the absence of signal processing are shown by filled circles and the thin continuous line. Mean scores after processing using the IBM are shown by filled squares, while mean word scores obtained after processing using the BM are shown by filled diamonds. At all SNRs the increases in word scores obtained by signal processing using either the IBM or BM are statistically significant ($p < 0.05$).

As previously noted, the IBM produced large increases in mean word scores when the unprocessed score was <50% (i.e., 30.7% at an SNR of -8 dB, and 27.2% at -6 dB). The increases in scores at SNRs of -4 and -2 dB were also substantial, 14.4% and 10.6% respectively, and can be seen to meet the targets suggested for our algorithms in Figure 12. Thus our IBM algorithm is judged to provide substantial improvements in speech intelligibility in the industrial-like noise and so is immediately applicable to situations in which speech and noise are available separately.

The BM algorithm produces consistent increases in mean word scores at all SNRs ranging from 4.8% to 7%. While, as expected, the increases are smaller than those produced by the IBM, they nevertheless confirm the potential for improving speech intelligibility by our signal processing during face-to-face communication. Indeed, the increases in word scores can be seen to come close to meeting the targets suggested for our algorithms in Figure 12 at word scores greater than ~60%.

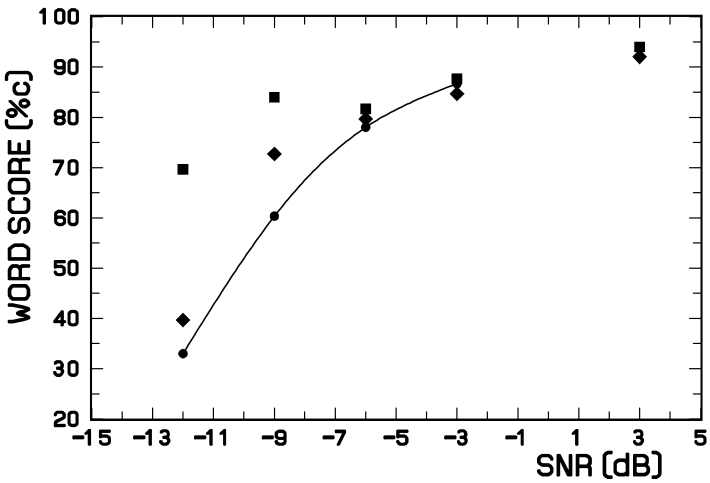


Figure 17: Mean word scores (% correct) at various speech SNRs in continuous miner noise. Scores when wearing a passive HPD are shown by filled circles and thin line. Word scores after processing using 24-subband algorithms are shown by filled diamonds for a BM, and filled squares for an IBM.

4.6.5 Performance of 24-Subband Binary Masking Algorithms in Mine Machine Noises

Listening tests were undertaken in the simulated noise of a continuous miner or roof bolter by fourteen subjects who possessed normal hearing (for continuous miner noise, mean age 26.5 years, range 21 - 44 years; and for rock bolter noise, mean age 28.3 years, range 20 - 44 years). The results are shown in Figures 17 and 18, respectively. In these diagrams mean word scores at various speech SNRs in the mine machine noises are shown as experienced by a miner wearing a passive HPD (filled circles) and when listening to our BM algorithm (filled diamonds) or IBM algorithm (filled squares). Results are shown for a wide range of SNRs (from -12 to +3 dB) and word scores before signal processing of from 33% to 86.7%.

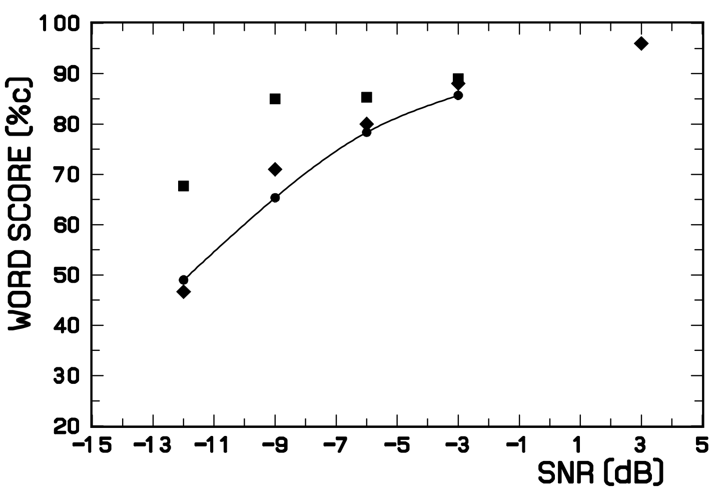


Figure 18: Mean word scores (% correct) at various speech SNRs in rock bolter noise. Scores when wearing a passive HPD are shown by filled circles and thin line. Word scores after processing using 24-subband algorithms are shown by filled diamonds for a BM, and filled squares for an IBM.

Compared to when wearing the passive HPD, the large increases in mean word score obtained by the IBM algorithm at low SNRs in industrial-like noise were again evident in continuous miner noise: in this case the increase was 36.7% at an SNR of -12 dB, and 23.7% at an SNR of -9 dB (see squares in Figure 17). Substantial, though not as large, increases in mean word scores were also obtained by this algorithm in rock bolter noise at low SNRs: in this case 18.7% at an SNR of -12 dB, and 19.7% at an SNR of -9 dB (see squares in Figure 18). All these increases in word scores were statistically significant ($p < 0.05$). The increase in mean word score in rock bolter noise was also statistically significant when the SNR was -6 dB and the word score when wearing the HPD was 78%. However, this was not the case for other SNRs at which there was a comparable word score when wearing an HPD and there was no signal processing (i.e., SNRs of -3 for both noises, and -6 dB for the continuous miner noise). As with speech intelligibility in the industrial-like noise, there is less need for the algorithm to increase word scores at these SNRs as they are already high without signal processing (viz., 86 - 87% at an SNR of - 3 dB, and 78% at -6 dB). The performance of the IBM algorithm was also evaluated in the absence of noise to establish whether the signal processing, including switching on and off of the mask, created sufficient distortion or artificial sounds to influence the word score. A mean word score of 94% was obtained under this condition, which was not statistically different from that obtained in the absence of signal processing (99.7%).

Overall, the performance of the IBM is considered to meet the targets suggested in Figure 12 for our algorithms when operating in continuous miner and rock bolter noise. Hence, as already noted when processing the industrial-like noise, this algorithm is judged to provide sufficient improvement in speech intelligibility to be immediately applicable to situations in which speech and noise are available separately.

The BM algorithm produced similar increases in mean word scores in listening tests using the industrial-like noise when the unprocessed scores ranged from ~40% to ~75% (i.e., increases of from 5% to 7%, see Figure 16). Reference to Figures 17 and 18 reveals that this pattern is less evident when comparing word scores obtained using the algorithm with those obtained wearing a simulated passive HPD and no signal processing. As with the performance of the BM in industrial-like noise, the changes in word scores are much less than those obtained when the IBM algorithm processed mine machine noises. For the BM algorithm the increases in mean word scores in continuous miner noise of 12.3% and 6.7% at SNRs of -9 and -12 dB, respectively, were statistically significant (see diamonds in Figure 17), while none were statistically significant when listening in rock bolter noise (see diamonds in Figure 18). The performance of the algorithm was also evaluated in the absence of noise for the reasons stated above. In this case mean word scores of 92 and 96% were obtained in two experiments. The former was statistically significantly different from that obtained when wearing the passive HPD in the absence of signal processing (99.7%), indicating a small reduction in word score was associated with production of the mask. There was no statistically significant difference in word score from that obtained in the second experiment when wearing the passive HPD in the absence of signal processing (99.3%).

The generally modest improvements in word scores produced by the BM algorithm and lack of statistically significant improvements in rock bolter noise, together with a possible reduction in word score in the absence of noise, are thought to be related to the formulation of the magnitude ratio. An attempt is being made to develop an improved mask after the completion of this study (see Appendix).

4.6.6 Performance of 24-Subband Binary Masking Algorithms in Intermittent Noise

In response to the mission statement the performance of the IBM and BM algorithms was also evaluated in intermittent noise. The noise was created in the following way in order to enable quantitative evaluation by listening tests employing the MRT. A 25-trial audio file was constructed consisting of:

Trials #1 - #10: Speech in noise at SNR = -8dB
Trials #11 - #15: Speech in noise at SNR = -2dB
Trials #16 - #20: Speech in noise at SNR = -8dB
Trials #21 - #25: Speech in noise at SNR = -2dB

Hence, in this test there were 15 trials in which noise dominated (trials #1 - #10, and #16 - #20) and 10 trials in which there was much less noise, perhaps modeling the background noise in a mine, that would have much less effect on intelligibility (trials #11 - #15, and #21 - #25).

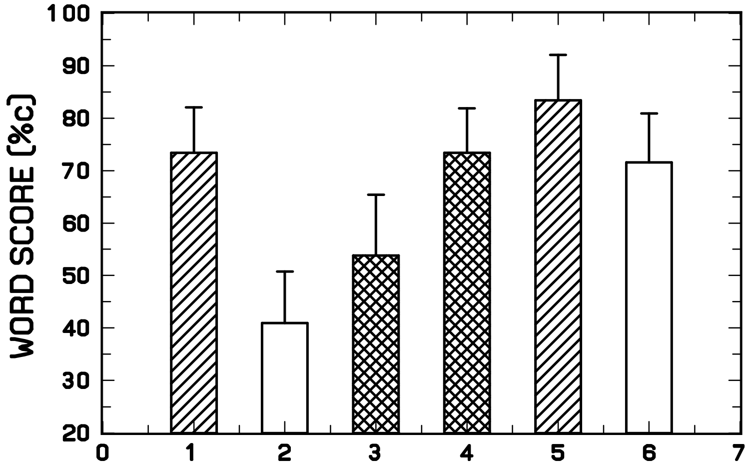


Figure 19: Mean word scores and SDs (% correct) for speech in industrial-like noise. Continuous noise unprocessed at SNR of -2 dB (column 1, shaded) and SNR of -8 dB (column 2); Intermittent noise at SNR -8 dB to -2 dB unprocessed (column 3, cross hatched) and processed by IBM (column 4, cross hatched); continuous noise processed by IBM at SNR of -2 dB (column 5, shaded) and SNR of -8 dB (column 6). For more information, see text.

Listening tests were undertaken in the simulated industrial-like noise by thirteen subjects who possessed normal hearing (mean age 26.5 years, range 20 - 44). The results are shown in Figures 19 and 20 for processing by algorithms containing an IBM and BM, respectively. In these bar graphs mean word scores (% correct) with SDs are shown both with and without signal processing. The performance of the algorithms in intermittent noise, which evaluates their capacity to improve speech intelligibility when sounds are changing in intensity, can be seen by comparing columns 3 and 4 in both diagrams (cross hatched shading). The former gives the mean word score before signal processing when the speech SNR is switching between -2 and -8 dB and the latter the word score after processing this intermittent noise by an algorithm containing an IBM (Figure 19) or a BM (Figure 20). It can be seen from the bar graphs that both algorithms can respond to the changing sound intensity and improve intelligibility with the IBM increasing the word score by 19.6% and the BM by 6.3%.

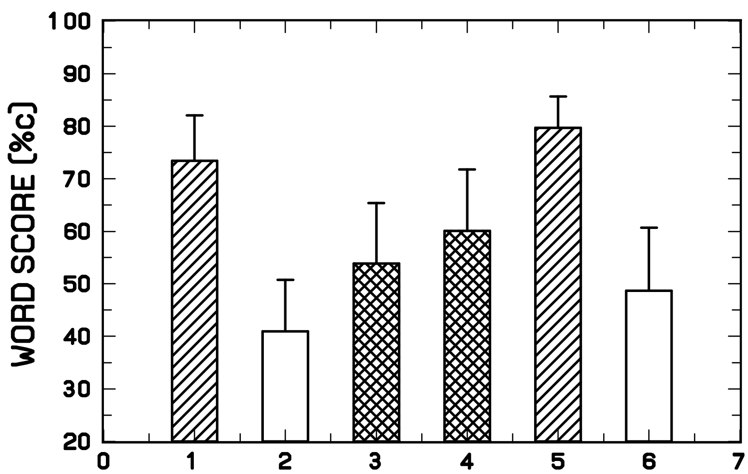


Figure 20: Mean word scores and SDs (% correct) for speech in industrial-like noise. Continuous noise unprocessed at SNR of -2 dB (column 1, shaded) and SNR of -8 dB (column 2); Intermittent noise at SNR -8 dB to -2 dB unprocessed (column 3, cross hatched) and processed by BM (column 4, cross hatched); continuous noise processed by BM at SNR of -2 dB (column 5, shaded) and SNR of -8 dB (column 6). For more information, see text.

The increase in mean word score produced by the IBM was statistically significant but this was not the case for the BM, where p = 0.1.

The other information in the bar graphs of Figures 19 and 20 enables these improvements in intelligibility in intermittent noise to be put in perspective. The word scores in columns 1 and 2 are for unprocessed speech in continuous noise at an SNR of -2 dB (shaded) and at an SNR of -8 dB, respectively. The word scores in columns 5 and 6 are for speech in continuous noise at an SNR of -2 dB (shaded) and at an SNR of -8 dB after processing by either the IBM algorithm (Figure 19) or the BM algorithm (Figure 20), respectively.

Reference to Figure 19 reveals the algorithm containing an IBM increased the word score in intermittent noise to equal that of the unprocessed continuous noise at an SNR of -2 dB (73.4% - compare columns #4 and #1). This performance is not as good as when the algorithm processed a continuous noise at an SNR of -2 dB (83.4% - column #5). Also, the word score obtained in intermittent noise when processed by the IBM algorithm is only slightly increased over that obtained when processing the continuous noise at an SNR of -8 dB (73.4% versus 71.6% - compare columns #4 and #6).

The performance of the algorithm containing the BM is somewhat different. Reference to Figure 20 shows the word score in intermittent noise when processed by the algorithm is less than that of the unprocessed continuous noise at an SNR of -2 dB (60.1% versus 73.4% - compare columns #4 and #1). The performance is also not as good as when the algorithm processed a continuous noise at an SNR of -2 dB (79.7% - column #5), but the word score in intermittent noise is a substantial improvement over that obtained when the algorithm processed continuous noise at an SNR of -8 dB (60.1% versus 48.7% - compare columns #4 and #6).

The difference in the performance of the two algorithms in intermittent noise is influenced by the ability of the IBM to improve speech intelligibility more than the BM in continuous noise at low SNRs, an observation that has been made previously and can be seen in the results of Figures 16 - 18.

5.0 Technology Capability Assessment and Readiness Assessment

Comments on the capability of the technology have been made in the previous section and are discussed further here. While the focus of algorithm development during the original project for the Alpha Foundation was restricted to face-to-face speech communication between persons wearing a hearing protector containing electronics or an eHPD, a second scenario is recognized here. This concerns the related situation wherein the talker's speech is intelligible and is available separately from the environmental noise as a second input for the control signal of the algorithm (see Figure 1). The second scenario may have application to mining when speech from a remote talker is transmitted over a wireless or wired link to a listener wearing a communication headset or eHPD. This scenario may be satisfied by an algorithm containing an IBM while the first requires an algorithm containing either a DM or a BM.

In this study, the performance of algorithms involving direct modulation (DM) and binary masking both with and without access to speech and noise separately, IBM and BM respectively, has been established in different noises and at different SNRs. The noises possessed a range of frequency spectra that either decreased, increased or remained unchanged in sound pressure level at frequencies below that of the maximum sound pressure level of speech (see Figure 4). At higher frequencies, all noises decreased in sound pressure level with increasing frequency at about the same rate as that of speech. It is thus believed that the performance of the algorithms has been determined for a range of noises typical of those expected to be found in mines. Additionally, the performance of algorithms employing binary masking has been determined in intermittent noise.

The use of different SNRs implies that listening tests were performed at different sound levels. Moreover, the conduct of listening tests in three different experimental settings - one conducted under controlled conditions within a controlled environment supervised by trained audiologists, a second conducted in an audiometric room in our laboratory where subjects themselves fitted their earphones, and a third commonly conducted at home where subjects chose the headphones or earphones to wear and the sound level for the tests - introduced an uncontrolled range of sound levels into each test. Thus we believe we have fully addressed the requirements of the mission statement.

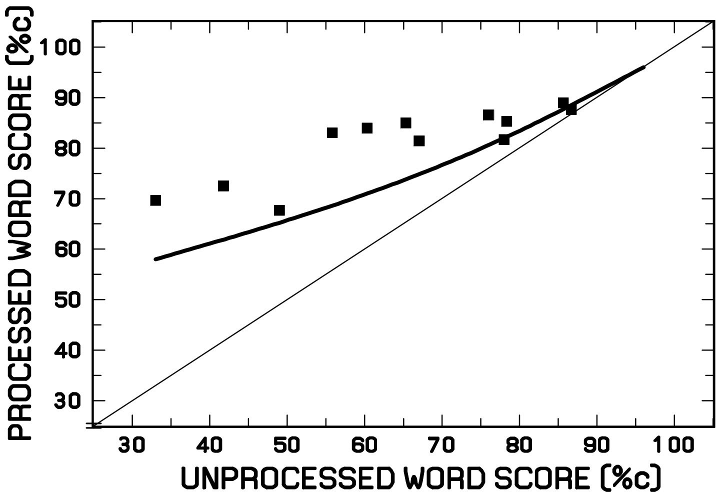


Figure 21: Mean word scores (% correct) for the same MRT first unprocessed and then processed by an IBM algorithm, shown by squares, for all noises and SNRs. Minimum target increase in word scores after processing designed to improve speech intelligibility is suggested by the thick line. Thin line indicates zero increase in word score achieved by signal processing (for further explanation, see text)

Under all the operating conditions imposed by the listening tests our 24-subband IBM algorithm provides improvements in speech intelligibility, which increase as the listening conditions become more challenging (i.e. large negative SNRs), as is desired. This is most easily confirmed by reference to Figure 21. In this diagram, mean word scores are expressed as the percentage of words correctly identified (%c), as before. Those obtained without signal processing, including the sounds heard when wearing a simulated passive HPD, are plotted on the abscissa and those after processing by the IBM algorithm on the ordinate. In this way the mean word scores obtained listening to the same audio file comprising one test of 25 trials can be represented in Figure 21 by a data point (shown as a filled square). Thus, if the mean word score recorded by subjects listening to unprocessed sounds is identical to that recorded when listening to the same test processed by our algorithm, then the filled square will fall on the thin line bisecting the graph diagonally (from coordinates 25,25 to 105,105). An improvement in intelligibility obtained by the signal processing will result in the square lying above the thin line. The minimum target increase in word scores after processing designed to improve speech intelligibility suggested in Figure 12 is shown here by the thick line. The importance of this presentation of the results lies in it capturing *all* listening conditions to *all* noises evaluated. Thus algorithms that produce word scores after signal processing on or above the thick line are judged to fulfill all requirements for improving speech intelligibility irrespective of the listening conditions - that is, listening in noise with little or substantial low frequencies, noise that is louder than the speech or not, and noise that changes in intensity.

In can be seen from Figure 21 that the performance of the IBM meets or exceeds the minimum target suggested in Figure 12 at all words scores recorded when speech in noise was not processed by our algorithm (from 33% to 87%). This large range of mean word scores was obtained in listening tests involving three different noises and seven different SNRs. The word scores recorded by individuals in tests without signal processing ranged from 8.8% to 95.2%, and from 53.2% to 100% after processing by the IBM. In consequence, we believe this algorithm has demonstrated the functional capability for in-service operational application to situations in which speech and noise are available separately.

The DM and BM algorithms produced smaller improvements in speech intelligibility than the IBM. The performance of these algorithms is presented in the same format as that for the IBM in Figure 22 for the BM algorithm and Figure 23 for the DM algorithm (see next page). Listening tests involved three different noises and seven different SNRs for the BM algorithm and two noises and three SNRs for the DM algorithm.

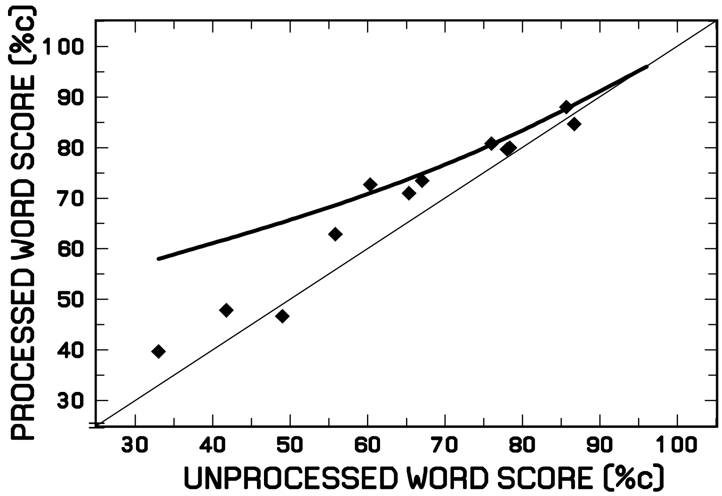


Figure 22: Mean word scores (% correct) for the same MRT first unprocessed and then processed by a BM algorithm, shown by diamonds, for all noises and SNRs. Minimum target increase in word scores after processing designed to improve speech intelligibility is suggested by the thick line. Thin line indicates zero increase in word score achieved by signal processing (for further explanation, see text)

Reference to Figure 22 reveals that the BM algorithm does achieve the minimum target performance suggested in Figure 12 for unprocessed speech in noise at word scores greater than ~60%. However, the target word score is not achieved at smaller initial word scores, though comparatively small improvements in intelligibility were obtained at three of the four word scores that were less than 60%. The word scores recorded by individuals in tests without signal processing ranged from 8.8% to 95.2% and from 34.7% to 100% after processing by the BM.

Reference to Figure 23 reveals the performance of the DM algorithm is similar to that of the BM in that it does achieve the minimum target suggested in Figure 12 for unprocessed speech in noise at word scores greater than ~67%. However, the target word score is not achieved at

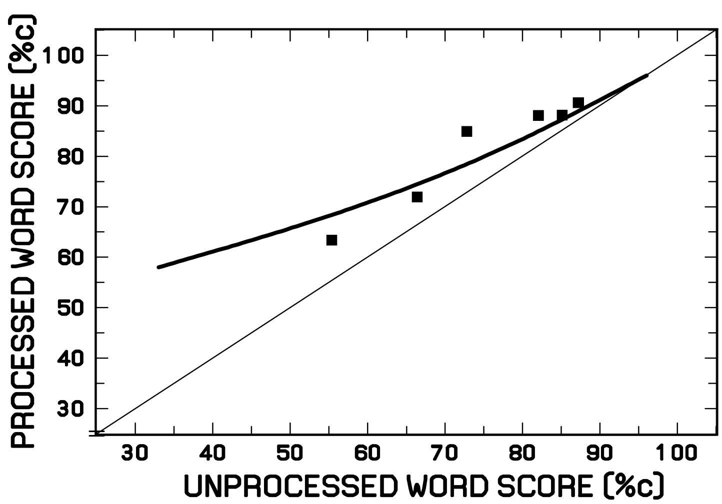


Figure 23: Mean word scores (% correct) for the same MRT first unprocessed and then processed by a DM algorithm, shown by squares, for all noises and SNRs. Minimum target increase in word scores after processing designed to improve speech intelligibility is suggested by the thick line. Thin line indicates zero increase in word score achieved by signal processing (for further explanation, see text)

smaller initial word scores. For individuals, the word scores in tests without signal processing ranged from 42.4% to 100% and from 47.2% to 100% after processing by the DM.

While smaller increases in word scores were expected for both algorithms, they nevertheless confirm the potential for improving speech intelligibility by our signal processing during face-to-face communication, when speech and noise are intermixed and never available separately. Implementation of either of these algorithms in an eHPD would provide modest improvements in intelligibility under almost all listening conditions, but both would benefit from further refinement. Based on the work reported here and the algorithms evaluated in our previous study for the Alpha Foundation, it is not clear how the DM algorithm could be modified to improve its performance. However, the limitations of the BM algorithm are thought to be related to the formulation of the magnitude ratio, and the potential increase in speech intelligibility achievable, in principle, is shown by the performance of the IBM algorithm. An attempt is being made to develop an improved mask after the completion of this study, and progress to date is described in an Appendix to this report.

Even after an algorithm is developed that can substantially increase the intelligibility of speech in a noisy environment, it must be transferred to electronics capable of microminiaturization. The computational complexity of 24-subband IBM and BM algorithms will require careful implementation within a small, lightweight package to function effectively throughout a work shift and be worn as part of a miner's equipment or attached to, or integrated into, a miner's helmet. Ultra low-powered digital signal processors (DSPs) or field-programmable gate arrays (FPGAs) will be required for a body-worn or helmet-mounted device and will need to be identified for this application. While the necessary performance can always be obtained by employing a second DSP, an FPGA may be more suited to this application in view of the amount of parallel processing (i.e., 24 parallel channels). Coding the device selected with the algorithm to provide the desired performance will require expertise and time.

We judge the Technology Readiness Level of the proof-of-concept to be level TRL 3 ("Analytical and experimental critical function and/or characteristic proof of concept"/NASA usage and "Experimental proof of concept"/European Union usage) of the current nine-unit scale (see https://en.wikipedia.org/wiki/Technology_readiness_level).

6.0 Publication Record and Dissemination Efforts

No presentations or publications have so far resulted from this work. The accomplishments are described in the previous sections of this report. A dissemination plan is not applicable at this time. A patent application based on the binary masking algorithms is anticipated.

7.0 Appendices

7.1 References

B. W. Anderson and G. R. Garinther, "Effects of active noise reduction in armor crew headsets," *Audio Effectiveness in Aviation,* AGARD-CP-596, pp. 20-1 - 20-6, (1997).

Anon., "Log loader runs over worker at stave mill," KY FACE Report 94KY161 (1995).

Anon., "NIOSH Criteria for a Recommended Standard: Occupational Exposure to Noise," DHHS (NIOSH) Publication 98-126, National Institute for Occupational Safety and Health, Cincinnati (1998).

ANSI S1.1-2004 (R2009), "Specifications for octave-band and fractional octave-band analog and digital filters," American National Standards Institute, New York (2004).

ANSI S3.1-1999 (R2003), "Maximum permissible ambient noise for audiometric test rooms," American National Standards Institute, New York (2003).

ANSI S3.2-1989 (R2009), "Method for measuring the intelligibility of speech over communication channels," American National Standards Institute, New York (2009).

ANSI S3.6-1996, "American National Standard specification for audiometers," American National Standards Institute, New York (1996).

M. C. Anzalone, L. Calandruccio, K. A. Doherty and L. H. Carney, "Determination of the potential benefit of time-frequency gain manipulation," Ear Hear. **27**, 480-492 (2006).

F. Apoux, N. Tribut, X. Debruille and C. Lorenzi, "Identification of envelope-expanded sentences in normal-hearing and hearing impaired listeners," Hear. Res. **189**, 13-24 (2004).

K. Arehart, P. Souza, J. Kates, T. Lunner and M. Pedersen, "Relationships among signal fidelity, hearing loss, and working memory for digital noise suppression," Ear Hear. **36**, 505-516 (2015).

A. A. Azman and R. L. Hudak (2011), "An evaluation of sound restoration hearing protection devices and audibility issues in mining," Noise Control Eng. J. **59**, 622-630.

D. R. Babich and E. R. Bauer, "Summary of longwall and continuous miner section noise studies in underground coal mines," Min. Eng. **58**, 41-46 (2006).

M. Bland, *An introduction to medical statistics*, Oxford University Press, Oxford U.K. (4th edition, 2015).

T. G. Bobick and D. A. Giardino, "The noise environment of the underground coal mine," Informational Report 1034, Mining Enforcement and Safety Administration, Washington (c1976), pp. 1-26.

A.J. Brammer, G. Yu, E. R. Bernstein, M. G. Cherniack, D. R. Peterson, and J. B. Tufts, "Understanding speech when wearing communication headsets and hearing protectors with subband processing," J. Acoust. Soc Am. **136**, 671-681 (2014).

D. S. Brungart, P. S. Chang, B. D. Simpson, and DeLiang Wang, "Isolating the energetic components of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007-4018 (2006).

H. E. Camargo, A. S. Azman and L. Alcorn, "Development of noise controls for longwall shearer cutting drums," Noise Control Eng. J. **64**, 573-585 (2016).

K. Cardosi, P. Falzarano and P. Han, "*Pilot-controller communication errors: An analysis of aviation safety reporting (ASRS) system reports*". DOT/FAA/AR-98/17. U.S. Department of Transportation, Office of Aviation Research, Washington (1998).

S.-W. Choi, C. Peek-Asa, N.L. Sprince, R.H. Rautiainen, K.J. Donham, G.A. Flamme, P.S. Whotten and C. Zwerling, "Hearing loss as a risk factor for agricultural injuries," Am. J. Ind. Med. **48**: 293-301 (2005).

K. Chung, "Effective compression and noise reduction configurations for application to hearing protectors," J. Acoust. Soc. Am. **121**, 1090-1101 (2007).

K. Chung, J. Tufts and L. Nelson, "Modulation-based digital noise reduction for application to hearing protectors to reduce noise and maintain intelligibility," Human Factors **51**, 78-89 (2009).

P. Clarkson and S. F. Bahgat, "Envelope expansion methods for speech enhancement," J. Acoust. Soc. Am. **89**, 1378-1382 (1991).

J. W. Collins, G. S. Smith, S. P. Baker and M. Warner, "Injuries related to forklifts and other powered industrial vehicles in automotive manufacturing," Am. J. Ind. Med. **36**, 513-521 (1999).

T. Dolan and D. O'Loughlin, "Amplified earmuffs: Impact on speech intelligibility in industrial noise for listeners with hearing loss," Am. J. Audiol. **14**, 80-85 (2005).
R. Drullman, J. M. Festen and R. Plomp, "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053-1064 (1994).
J. R. Dubno, A. R. Horowitz and J. B. Ahlstrom, "Word recognition in noise at higher-than-normal levels: Decreases in scores and increases in masking," J. Acoust. Soc. Am. **118**, 914-922 (2005).
ECMA TR/105. "A shaped noise file representative of speech," European Computer Manufacturers Association, Geneva (2012), pp. 1-13.
A. S. House, C. W. Williams, M. H. L. Hecker and K. D. Kryter, "Articulation testing methods: Consonant differentiation with a closed-response set," J. Acoust. Soc. Am. **37**, 158-166 (1965).
G. J. Joy and P. J. Middendorf, "Noise exposure and hearing conservation in U. S. coal mines - A surveillance report," J. Occup. Environ. Hygiene **4**, 26-35 (2007).
G. Kim, Y. Lu, Y. Hu and P. C. Loizou, "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," J. Acoust. Soc. Am. **126**, 1486-1494 (2009).
U. Kjems, J. B. Boldt, M. S. Pedersen, T. Lunner and DeLiang Wang, "Role of mask pattern in intelligibility of ideal binary-masked noisy speech," J. Acoust. Soc. Am. **126**, 1415-1426 (2009).
T. LaTourette, D. J. Peterson, J. T. Bartis, B. A. Jackson and A. Houser, *Protecting Emergency Responders: Community Views of Safety and Health Risks and Personal Protection Needs*, Vol. 2, Rand Science and Technology Institute (2003).
T. Langhans and H. Strube, "Speech enhancement by nonlinear multiband envelope filtering. In: *ICASSP '82, IEEE International Conference on Acoustics, Speech and Signal Processing,* edited by C. Gueguen, Institute of Electrical and Electronic Engineers, New York (1982), pp. 156-159.
N. Lezzoum, G. Gagnon and J. Voix, "Noise reduction of speech signals using time-varying and multi-band adaptive gain contol for smart digital hearing protectors," Applied Acoustics **109**, 37-43 (2016).
N. Li and P. C. Loizou, " Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," J. Acoust. Soc. Am. **123**, 1673-1682 (2008).
C. Lorenzi, F. Berthommier, F. Apoux and N. Bacri, "Effects of envelope expansion on speech recognition," Hear. Res. **136**, 131-138 (1999).
R. L. McKinley, V.S. Bjorn and J.A. Hall, "Improved hearing protection for aviation personnel," NATO Technical Report RTO-MP-HFM-123, Neuilly-Sur-Seine, France, pp. 13-1 – 13-12 (2005).
Mines Safety Health Administration web site www.msha.gov/STATS/PART50/WQ/2004/WQ045T03.asp
B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, New York (6th edition, 2013).
T. C. Morata, A.C. Fiorini, F.M. Fischer, E.F. Krieg, L. Gozzoli and S. Colacioppo, "Factors affecting the use of hearing protectors in a population of printing workers," Noise and Health **4**, 25-32 (2001).
T. C. Morata, C. L. Themann, R. F. Randolph, B. L. Verbsky, D. C. Byrne and E. R. Reeves, "Working in noise with a hearing loss: Perceptions from workers, supervisors, and hearing conservation program managers," Ear Hear **26**, 529-545 (2005).
L. Murray-Johnson, K. Witte, D. Patel, V. Orrego, C. Zuckerman, A.M. Maxfield and E.D. Thimons "Using the extended parallel process model to prevent noise-induced hearing loss among coal miners in Appalachia," Health Education Behavior **31**, 741-755 (2004).
D. S. Patel, K. Witte, C. Zukerman, L. Murray-Johnson, V. Orrego, A.M. Maxfield, S. Meadows-Hogan, J. Tisdale and E.D. Thimons, "Understanding barriers to preventive health actions for occupational noise-induced hearing loss," J. Health Commun. **6**, 155-168 (2001).
P. Plyler and M. Klumpp, "Communication in noise with acoustic and electronic hearing protection devices," J. Am. Acad. Audiol.**14**, 260-268 (2003).
R. S. Schlauch and P. Nelson, "Pure-tone evaluation," in *Handbook of Clinical Audiology*, edited by J. Katz, L. Medwetsky, R. Burkard and L. Hood, Lippincott Williams & Wilkins, Baltimore (6th edition, 2009), pp. 30-49.

A. Suter, "The handicap resulting from noise-induced hearing loss," *Best Practices in Hearing Loss Prevention*, DHHS (NIOSH) Pub. 2001-157, Department of Health and Human Services, Washington, pp. 3-8 (2001).
A. H. Suter, Communication and Job Performance in Noise: A Review, Monograph 28, American Speech-Language-Hearing Association, Rockville (1992).
M. L. Szary, Y. P. Chugh and J. Hirschi, "Noise variability in underground room and pillar coal mines," Int. J. Occup. Safety & Ergonom. **17**, 301-308 (2011).
S. Tak, R. R. Davis and G. M. Calvert, "Exposure to hazardous workplace noise and use of hearing protection devices among US workers - NHANES, 1999-3004," Am. J. Industr. Med. 52, 358-371 (2009).
J. B. Tufts, M. A. Hamilton, A. J. Ucci and J. Rubas, "Evaluation by industrial workers of passive and level-dependent hearing protectors," Noise & Health, **13**, 26-36 (2011).
R. A. van Buuren, J. M. Festen and T. Houtgast, "Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality," J. Acoust. Soc. Am. **105**, 2903-2913 (1999).
M. van Charante, A.W. Mulder and P.G.H. Mulder, "Perceptual acuity and the risk of industrial accidents," Am. J. Epidemiol. **131,** 652-663 (1990).
R. M. Warren, K. R. Riener, J. A. Bashford and B. S. Brubaker, "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," Percept. Psychophysics **57**, 175-182 (1995).
A. Wiinberg, J. Zaar and T. Dau, "Effects of expanding envelope fluctuations on consonant perception in hearing-impaired listeners," Trends in Hearing **22**, 1-12 (2018).
P. Wilkins and A. M. Martin, "Hearing protection and warning sounds in industry – A review," Applied Acoustics **21**, 267-293 (1987).
K. K. Wójcicki and P. C. Loizou, "Channel selection in the modulation domain for improved speech intelligibility in noise," J. Acoust. Soc. Am. **131**, 2904-2913 (2012).

7.2 Hearing Function Questionnaire

Please circle your response to the following questions.

1. Does a hearing problem cause you to feel embarrassed when meeting new people?
   YES
   SOMETIMES
   NO

2. Does a hearing problem cause you to feel frustrated when talking to members of your family?
   YES
   SOMETIMES
   NO

3. Do you have difficulty hearing when someone speaks in a whisper?
   YES
   SOMETIMES
   NO

4. Do you feel handicapped by a hearing problem?
   YES
   SOMETIMES
   NO

5. Does a hearing problem cause you difficulty when visiting friends, relatives, or neighbors?
   YES
   SOMETIMES
   NO

6. Does a hearing problem cause you to attend religious services less often than you would like?
   YES
   SOMETIMES
   NO

7. Does a hearing problem cause you to have arguments with family members?
   YES
   SOMETIMES
   NO

8. Does a hearing problem cause you difficulty when listening to TV or radio?
   YES
   SOMETIMES
   NO

9. Do you feel that any difficulty with your hearing limits or hampers your personal or social life?
   YES
   SOMETIMES
   NO

10. Does a hearing problem cause you difficulty when in a restaurant with relatives or friends?
    YES
    SOMETIMES
    NO

SCORE __________ [Examiner score: No - 0; Sometimes - 1; Yes - 2]

7.3 Postscript - Improving Magnitude Ratio for Detecting Speech in Noise

Following completion of the study for the Alpha Foundation on "Improving Communication in Noise for Miners Wearing Hearing Protection: Algorithms for Mine Machinery Noise" it was evident that improvements were required to the magnitude ratio in order to improve the detection of speech in noise. An attempt is being made to derive the magnitude ratio from the ratio of envelopes representing an estimate of the speech to an estimate of the noise at *all* modulation frequencies in a subband.

The initial results are shown for one subband in Figure A1 below. In this diagram the performance of the original magnitude ratio, which was used to obtain the results in this report, is shown to the left (Figure A1A), and the revised magnitude ratio is shown to the right (Figure A1B). Below each are the corresponding time-aligned signals for the IBM, which are the same in both cases.

Close inspection of the magnitude ratio time histories reveals that the amplitude of the revised magnitude ratio is much greater than the original version (i.e., compare scales of top waveforms under "Binary Mask" in Figure A1B with Figure A1A). In fact, while the peaks of the new metric are approximately twice the amplitude of those of the original magnitude ratio, the baseline of the background "noise" is close to zero while that of the original metric is close to

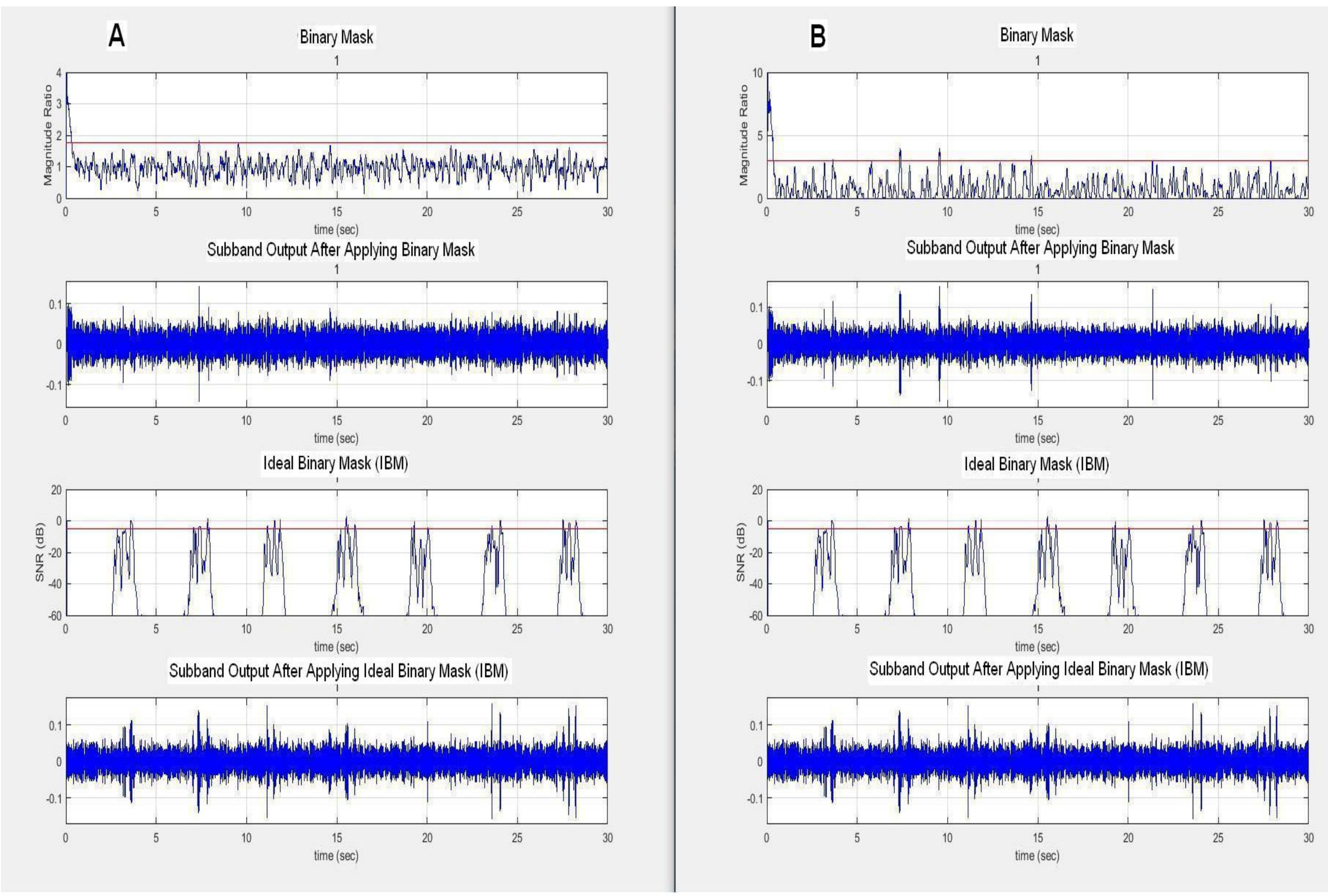


Figure A1: Time-aligned time histories for one subband showing binary mask (BM) and ideal binary mask (IBM) processing of speech in noise. A - Original BM used to compute results in this report, B - revised BM. Mask waveforms are shown for BM and IBM with corresponding subband outputs.

unity. These observations suggest that the revised magnitude ratio will possess better resolution of speech in noise than the original version and hence improve detection of speech in noise. This belief may be confirmed by examining the subband output after applying the different binary masks.

If attention is turned to the lower parts of the diagrams, which are the same in Figures A1A and A1B, it can be seen that the IBM detects speech seven times during the time period shown, that is, each time the SNR exceeds the threshold of -5 dB (shown by the horizontal red line). Reference to the subband output after applying the IBM shows the sounds transmitted by the mask, which appear as seven peaks of varying intensity in a background of environmental noise (see bottom time histories in Figures A1A and A1B). That these peaks are in fact MRT sentences separated by ~3 s has been confirmed by listening tests.

If, now, the subband outputs of the two binary masks are compared, it can be seen that more of the speech sounds are transmitted by the new mask as evidenced by larger peak values or longer outputs (i.e., compare magnitudes and/or durations of subband outputs after applying the binary masks at around 3, 7,15 and 28 s). However, spurious signals are incorrectly identified as speech by the new mask at around 9 and 22 s (and also, though less definitively, by the original mask). The reason for the false positives is unclear, though they are most likely associated with details of the computation of the magnitude ratio and selection of the mask threshold. They will, of course, be the subject for future work involving further revision of the mask. Nevertheless, the inclusion of all subband modulation frequencies in the new mask and its improved resolution compared to the original mask are considered of great significance for improving speech intelligibility in noise by a BM algorithm.